Towards an Ecosystem for Semantics

Ansgar Scherp
OFFIS - Institute for Information Technology
Oldenburg, Germany
ascherp@ics.uci.edu

Ramesh Jain
University of California
Irvine, USA
jain@ics.uci.edu

ABSTRACT

Multimedia does not exhibit a unique semantics but multiple semantics that are influenced by many factors. Current approaches and systems lack from considering this problem in its entirety. What is needed is a holistic approach that describes and embraces the complex and challenging problem of multimedia semantics. Consequently, we are developing with the Semantics Ecosystem (SES) an approach that aims at providing a "framework" for solving this problem. The SES defines five types of semantics and their relationships to each other and to the real world. In the paper, we present the SES and show its application to the domain of authoring multimedia albums. We show the benefits of having such a theoretical framework to handle the semantics of multimedia albums. The SES allows us to better understand, describe, and communicate the many different factors that are influencing multimedia semantics.

Categories and Subject Descriptors

H.1.1 [**Information Systems**]: Systems and Information Theory—*information theory, value of information*; H.1.2 [**Information Systems**]: User/Machine Systems—*human factors, human information processing*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems; H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia

General Terms

Human Factors, Management, Design, Algorithms

Keywords

Semantics, Multimedia Semantics, Events, Ecosystem, Multimedia Authoring

1. INTRODUCTION

In recent years, it has become more and more clear that multimedia content does not have a pre-defined, unique semantics but inhibits multiple semantics [28]. This multiple semantics is influenced by many different factors like time and context. The management of the multiple semantics of multimedia content has become a crucial aspect in modern multimedia systems and gains more and more interest in research. However, current research in multimedia semantics and today's approaches and systems lack in considering the problem in its entirety. Today's research and systems are mostly focused on single aspects of multimedia semantics such as the change of semantics over time through user interaction with the system [10, 14].

What is missing is a comprehensive approach to describe and to embrace this very complex and challenging problem of multimedia semantics. We strongly believe that only by looking at the problem holistically, we will gain a better understanding of what multimedia semantics is and how we can deal with it efficiently and effectively. Consequently, we are developing the Semantics Ecosystem (SES) approach that aims at providing a "framework" for solving this problem. The SES defines five types of semantics. These are the Natural, Analytical, User, Expressive, and Emergent Semantics. The paper elaborates these five types of semantics, describes the relationships between the different semantics, and shows how they influence each other, the real world, and vice versa.

The SES has been primarily motivated by the problem of understanding the multiple factors that influence the semantics of multimedia presentations created in an authoring tool. So far, the semantics of multimedia presentations has been only implicit with the multimedia content. This implicit semantics shall be made explicit by extracting and expressing it in semantics derivation rules that are applied during the authoring process. We are currently focusing on Expressive Semantics as it allows for a comparable easy abstraction and expression in rules and thus is most promising. Early results are presented and the implementation of example rules for Expressive Semantics in a context-driven authoring tool for page-based multimedia content [17] is shown.

The remainder of the paper is organized as follows: In Section 2, we introduce our Semantics Ecosystem. The different kinds of semantics defined in this ecosystem, their relationships to each other, as well as their relationship to the real world are presented in Section 3. In Section 4, we describe how the Semantics Ecosystem helps us in understanding, defining, and communicating the semantics of multimedia albums. Example rules of Expressive Semantics for multime-

dia albums are identified and defined in Section 5. We also present how some other semantics namely User Semantics and Emergent Semantics affects the semantics of multimedia albums. In Section 6, we critically discuss the Semantics Ecosystem and point out the most important challenges that need to be solved, before we conclude the paper.

2. SEMANTICS ECOSYSTEM (SES)

For the definition of the Semantics Ecosystem (SES), we base on the existence of a Physical World and a Mental Model as depicted in Figure 1. The Physical World is the world around us and in which we live. It is the stones, trees, streets, houses, and coffee on our breakfast table [13]. All these objects are part of our natural environment and become deeply embedded in our thinking effortlessly. We start considering these objects and the symbols associated with them part of the nature. Natural semantics is associated with these. The Mental Model is our cognitive construction of the Physical World. It contains our very individual observations and experiences how the Physical World works and what the rules in this world are. Thus, the Mental Model is the result of our conscious experiences in the Physical World [13].
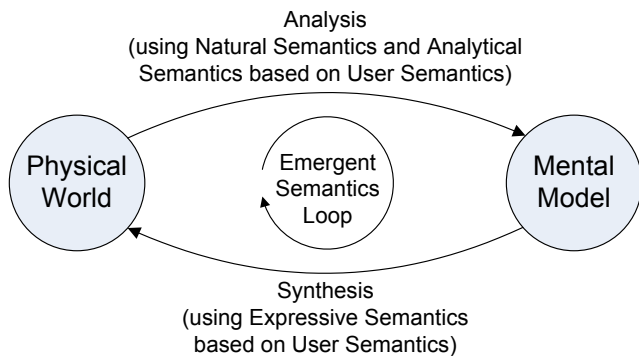
Analysis
(using Natural Semantics and Analytical
Semantics based on User Semantics)

Physical World — Emergent Semantics Loop — Mental Model

Synthesis
(using Expressive Semantics
based on User Semantics)

Figure 1: The Semantics Ecosystem

By *analysis*, we come from the Physical World to create our Mental Model. This analysis leverages the so-called Natural Semantics as well as Analytical Semantics. It means that we recognize, analyze, and interpret the non-living physical objects, the living things, and the events of our Physical World. We see, observe, and experience them. This analysis is not unbiased. Rather, it is strongly influenced by our very individual background, from our history, education, knowledge, experience, culture, interests, preferences, and many more. These different factors that influence the analysis are called the User Semantics.

With *synthesis*, our Mental Model influences and manipulates the Physical World. Influencing the Physical World is done by creating "products" based on our Mental World. These products can be scientific theories taught in universities using natural language, books written for the general public, photos, movies, or multimedia presentations created to convey a certain message. For synthesizing these products, we apply the so-called Expressive Semantics. It is the semantics we apply to express a certain message or convey a specific information. Again, synthesis is—like analysis—very much influenced by our individual User Semantics.

This means, the way we express a certain message depends on our very individual background.

The loop between the two worlds, between analysis and synthesis is the so-called Emergent Semantics. It is not about the semantics of the physical objects or media like photos or multimedia presentations but is about the change of our Mental Model over time through the change of User Semantics. This change happens because of our different analysis of the Physical World when our personal background, history, and so forth changes, i.e., when we are making new experiences in the Physical World.

All this is creating in a sense a notion of *personal ontology* that exists only in our head, as compared to a general ontology that is based on consensus knowledge used in communication. These experiences can and very likely will be influenced by (many) other persons' Mental Model products. For example, the movie created by a famous director we saw last night, the book read the other day, or the multimedia presentation we saw at work on our screen. Consequently, there is a multitude of different, very individual Emergent Semantics loops. All bound together by the one Physical World.

As a note, let us mention that the ecosystem could consider besides the Physical World and Mental Model also a third world that Karl Popper [24] introduced [13]. This third world is the world of the Products of Human Mind. Elements of this world are among others the theories, songs, and symphonies created from the human' thoughts ("stored" in the Mental Model or as he calls it the Mental World[1]), without instantiating them into a concrete product. In Popper's approach, these elements are considered abstract concepts. Though, they are considered real as the third world itself is also considered real. The abstract concepts can then be embodied in physical products of the Physical World like a CD of Beethoven's 9th Symphony or a book about Einstein's relativity theory. The three-world theory is not undisputed as Popper himself discusses in his work [13]. For our consideration of multimedia semantics, at least at this stage, it does not make a difference whether we base the ecosystem on a two-world or three-world approach. Thus, we do not follow this three-world approach and consider in our ecosystem only the Physical World and the Mental Model existing.

The presented Semantics Ecosystem is only a very abstract description of the different kinds of semantics involved. In fact, the relationships and interaction between the Physical World and the Mental Model by the different semantics involved is much more complicated and sophisticated than presented here in the diagram and its description. However, the presented model is already powerful enough to sufficiently identify and define the different kinds of semantics and to describe the relationships of them in our ecosystem. The different kinds of semantics and in particular their relationships are elaborated and discussed in detail in the following section.

---

[1]Note: Our thoughts and brain processes have less in common with the characteristics of the Physical World. Thus, we do not use the term Mental World in our ecosystem as we think that the thoughts and brain processes in our mind are better described as a Mental Model.

3. KINDS OF SEMANTICS AND THEIR RELATIONSHIPS

In the previous section, we introduced the Semantics Ecosystem. We mentioned a couple of different semantics involved in this ecosystem, the Natural Semantics, Analytical Semantics, User Semantics, Expressive Semantics, and Emergent Semantics. In this section, we introduce and elaborate these different semantics in detail.

3.1 Natural Semantics

With natural semantics, we understand the semantics of the non-living physical objects, living things, and events of our Physical World. Non-living physical objects are any items like rocks, streets, cars, signs, and so on. Living things or biological objects are humans and animals. An event is the change of relation between any item or biological object or the change of any characteristic attribute of these items and biological objects [25, 26]. The Natural Semantics of items, biological objects, and events evolved over the thousands of years of human development. We have learned it in our lives ever since we have been toddlers. Natural Semantics is the result of a natural language communication between humans and thus the one most commonly used.

Examples of Natural Semantics are the recognition of mother's face, a street sign, or a soccer goal (ball has crossed the white line), and realizing not to touch a hot stovetop. Other examples are our reaction when finding a $100 bill on the street or spraining one's toe [6]. Like all human' interpretation of things, Natural Semantics is subjective. It depends on the individual education, history, background, and many more, i.e., it depends on the User Semantics.

3.2 Analytical Semantics

Natural Semantics associates basic objects and actions with symbols. We use Analytical Semantics to understand more complex objects, concepts, and situations. The Analytical Semantics is applied to dismantle them, identify the individual parts, and look at the distinct parts. These parts are then interpreted by applying our Natural Semantics introduced above.

Analytical Semantics is used, e.g., for analyzing more complex Physical World events like the behavior of people at a demonstration or the defense of the soccer team in the first half. It also considers the analysis and interpretation of complex items such as an unknown space rock under a microscope or a person's x-ray that might show a tumor. Another example for Analytical Semantics is when we observe a car accident. We have seen a car coming from left driving much too fast. The car from the right tries to stop, breaks squeak. However, it is too late. The two cars crash into each other but luckily nobody gets hurt. We analyze the event by our Analytical Semantics and interpret all the details of the event. We observe and interpret them using Natural Semantics. We combine them together to get a picture of the entire event. By this, we are able to recognize that it is a car accident. We might start analyzing it in regard of some further questions like "How could that happen?" or "Who is at fault?".

As a consequence, we understand Analytical Semantics as implementing certain operators that allow us to combine in a specific way basic objects or concepts from Natural Semantics. With these operators at hand, we are able to understand the more complex objects, concepts, and situations

in our Physical World. To give a simple example, consider the situation where a child is learning about the concept of sunset. This concept seems quite simple to us and we probably would consider it as part of our Natural Semantics. However, for a child it is a complex situation that requires learning Analytical Semantics operators in order to identify and understand the concept. These operators include knowing that the sun is turning to red, close to the horizon, the clouds in the sky are illuminated red, and time is going closer to evening. By knowing these basic concepts, we are able to learn identifying a sunset.

The above example shows that there is a strong relationship between Natural Semantics and Analytical Semantics. However, it shows also that there is a shade of gray between Natural Semantics and Analytical Semantics. One could argue that Natural Semantics, the recognition of faces and street signs, is an instantaneous or spontaneous process. Whereas Analytical Semantics, we could argue is a conscious brain process that requires a much higher cognitive load. However, we see in many cases like the example above that Analytical Semantics may slowly—with practice—become Natural Semantics. As Natural Semantics is subjective, also Analytical Semantics is and very much depends on the User Semantics described next.

3.3 User Semantics

The User Semantics is a concept of the human's perception of the Physical World based on his or her personal background information. To be more precise, it is the perception of the items, biological objects, and events of the Physical World based on a multitude of very different aspects. Part of these aspects are the users' profile information that comprise among others the knowledge, preferences, interests, needs, and cultural background [5, 11, 8]. These aspects are typically subsumed under the notion of user modeling. User Semantics is also influenced by the context of the users such as the location, time, used end device, and their social situation such as whether the user is alone, with her friends, co-workers, or boss [21, 7, 22]. This is typically elaborated under the notion of context modeling, which became popular in the nineties with the emergence of small mobile devices like cell phones and personal digital assistants with plenty of sensors build in.

The aspects of both user modeling and context modeling have been elaborated in detail by Scherp [16, Sec. 6.2] from the perspective of authoring personalized multimedia content. In order to fully reflect the complexity of the different aspects, they may be represented as *personal ontologies*. This will allow clear and explicit representation of a user's specific semantics. It may be possible to represent relationships among the personal ontologies with other ontologies to clearly specify where the commonalities and differences arise.

The discussion shows, User Semantics is a complex concept in itself and very different for each individual person. It plays an important role for our interpretation of the Physical World by means of Natural Semantics and Analytical Semantics. Examples for User Semantics are street signs that are only typical for Australia. Thus, they are only immediately recognized in terms of Natural Semantics by people who where either born and raised there or have lived there for a longer period of time [4]. Another example is the number 4 that typically has no specific connotation for

people from Europe. However, it has a very specific meaning for people from Asia like China and Japan. Besides its meaning as number four it also has the connotation of death and misfortune (see, e. g., [27]).

3.4 Expressive Semantics

With Expressive Semantics, we consider how the products of the Physical World are created. We analyze, why they are created in that certain way and what is the intention of the creator in creating it like this. The product can be a gesture, a spoken sentence, or any kind of a non-living object created on the basis of one's Mental World like a book, CD, or multimedia presentation.

Examples of Expressive Semantics are when a person makes a questioning mien to express that he does not understand or cannot follow the current conversation. Another example is, when a homeowner puts a board on his land that shows "Too much traffic!" in order to complain about the traffic situation. In the media world, Expressive Semantics means for photos the way we take these photographs, e. g., the angle the photographer has chosen to capture the moment. In a movie, it is the narrative structure defined by, e. g., which clips are captured, in which order they are shown, how long the duration of each clip is, and what kind of background music is selected. For a detailed discussion on narrative structures for movies see, e. g., [2]. For a multimedia presentation, Expressive Semantics is reflected, e. g., in the media assets that are selected for the presentation and how they are organized in time and space.

If some of the media assets have been used earlier, their semantics might change with the new presentation as the media assets are used in a new contextual situation. This change of semantics over time and use is part of the Emergent Semantics. It is also important to note that Expressive Semantics is heavily influenced by our individual User Semantics. This means that the way we select clips for a movie, which media assets we select for a presentation, and how we arrange them in time and space is influenced by our education, background, history and others. We are creating the product in a specific way because of our personal background *how these should* be organized. The Expressive Semantics describes this way of *how to organize* in order to conveying a certain message to the audience. Thus, when doing so, we apply parts of our very individual background, i. e., our User Semantics.

Another example showing the strong relationship of Expressive Semantics with User Semantics is the depiction of a family in different cultural circles. Whereas an Indian family picture will show a hierarchy of people, the depiction of a family in China will show all persons being equally large. This Expressive Semantics reflects the general understanding of the relationship of the different family members in these two cultural circles.

Some traditional areas have developed rules and conventions for Expressive Semantics. For example, a very well developed area is video production. Movie and TV program producers have well defined sets of rules that they use to express importance of objects and events and for combining shots into scenes and then into stories [2, 1]. Painting also has their own rules for expressing relationships and importance of objects and events. Interestingly, different cultures use different Expressive Semantics and people from those cultures are trained to not only interpret that (*analysis*) but also to start using them in their own expressions (*synthesis*). Thus, a Emergent Semantics loop is established.

Finally, Expressive Semantics can also be understood as a kind of personalization. To achieve maximum success in conveying a certain message to our target audience, we use Expressive Semantics to adapt the Physical World product we are creating to our target group. When creating a slide show presentation for the next supervisory meeting in which one wants to convince their co-workers and boss, one is anticipating the target audience's User Semantics (i. e., the way the audience will very likely perceive the message) and adapt the presentation according to it. Thus, the presentation is very likely designed in a way that optimizes the chance that one's ideas get accepted. In another scenario, the teacher wants to explain a certain subject to his students. The teacher is preparing the lesson according to his knowledge of didactics, i. e., his background and Expressive Semantics rules of how to teach and according to this adapts the lesson to the knowledge level of the students.

3.5 Emergent Semantics

The term Emergent Semantics is used ambiguously in literature. For example, it is considered to describe browsing paths of web pages [10, 23, 9], user created relationships between images [14, 15], or derived from the authoring process of multimedia presentations [20]. In this work, we refer to the term Emergent Semantics as the change of semantics over time and use. It means that the individual's semantics and observation of a Physical World item, biological object, or event can and will change over time and will change trough the different contexts in which it is used. Emergent Semantics can be short-termed, i. e., a very short time spawn is considered. This may be only a couple of seconds up to some minutes. It can also be long-termed where potentially a very long time period is considered like a couple of years. However, the key to Emergent Semantics is the interaction of Expressive Semantics and Analytical Semantics. This interaction leads to a modification of User Semantics, i. e., the personal ontologies of the Mental Model and by this eventually the SES. This loop may sometimes happen very fast, but in other cases it may take longer durations.

The strong correlation of Emergent Semantics with the way how the products of the Physical World are created, i. e., how the media assets are used according to one's Expressive Semantics can be understood as the different contexts in which the media assets may be used over time. For example, the semantics of a photo used several times in a newspaper can change over a long period of time. The reason for this change of semantics can be the different articles in which the photo has been used. Although the articles have some common topic, they may present different aspects, focus on different matters, or reveal new information and facts about the subject.

However, Emergent Semantics can also appear by "just" making new experiences in the Physical World. As these experiences change the User Semantics, the perception of items, biological objects, and previous events might change. For example, a picture one took at a very exciting vacation ten years ago might nowadays not be considered so important anymore. The vacation's importance may have gradually decreased as the person has been to other places in the meantime or has discovered other challenges or interesting facets in her live.

An example of a more short-term Emergent Semantics appears when two or more people are discussing about a specific subject. By exchanging their opinions, concepts, and notions, they influence the User Semantics and thus the Mental Model of the others. Thus, the semantics of the discussed subject emerges through their talk.

3.6   Summary

In this section, we elaborated in detail the different kinds of semantics in our ecosystem. We presented these different kinds and discussed each of them. This discussion showed that there are plenty relationships between the different semantics. Thus, the semantics presented are not distinct, encapsulated parts in the ecosystem. Rather, they are factors that highly influence each other.

4.   PUTTING SES INTO PRACTICE

The development of the Semantics Ecosystem was initially motivated by a very practical problem we wanted to solve. The problem was to develop a system that allows to semantically-enrich multimedia albums during the authoring process of such albums. A multimedia album is a page-based multimedia presentation. It extends a traditional, printed photo book by the use of media types other than images and text like audio and video. It also has a temporal flow and navigational interaction by providing links. The difficulty is how to deal with the semantics of such a multimedia album. As more and more recognized by the multimedia community, there are many different factors that influence the semantics of multimedia. In order to be able to design and implement a system that allows for semantics derivation for these presentations, we need a good understanding of the factors influencing the multimedia content's semantics and thus the semantics derivation rules.

As a consequence, we started to investigate in the field of semantics and developed the Semantics Ecosystem. Having this ecosystem allows us to precisely describe what we are developing and designing in our system. Thus, we benefit from the fact that we have a theoretical "framework" for semantics and use this framework to describe our very practical problem of semantically-enhancing the semantics of multimedia albums. This allows us to better describe what is "semantically" happening with the multimedia album when creating it. In addition, being able to more precisely talk about the authored content's semantics enables us to better communicate our work with other researchers.

So far, we are focusing on the Expressive Semantics of multimedia albums. We chose Expressive Semantics as it allows for a comparable easy abstraction and expression in rules and thus is most promising. We identified a set of Expressive Semantics rules that allow for semantically-enriching an album. We implemented the rules in the context-driven multimedia authoring tool xSMART [17, 20] for page-based multimedia presentations. As the xSMART tool bases on an abstract software framework that allows for semantics derivation [20], the Expressive Semantics rules are implemented and integrated into our general authoring process for multimedia presentations provided by the framework [16, 19, 18]. The discovery of the Expressive Semantics rules and their integration into our xSMART tool is presented in the following section.

5.   SEMANTICS IN MULTIMEDIA ALBUMS

As we elaborated in Section 3.4, some areas like video production and painting have developed rules for Expressive Semantics. These rules apply for their particular, single medium. Now, we are considering Expressive Semantics of *multi*media content in the form of multimedia albums where disparate media with disparate characteristics are being used. In Section 5.1, we first present how we identified the Expressive Semantics rules for authoring multimedia albums. Example rules for Expressive Semantics are presented in Section 5.2. Further rules for Expressive Semantics are briefly described in Section 5.3. The extension to other kinds of semantics, i.e., User Semantics and Emergent Semantics is presented in Section 5.4.

5.1   Discovering Expressive Semantics Rules

The Expressive Semantics rules are discovered and discussed in collaboration with our project partner CeWe Color, Europe's leading photo finisher. We conducted an analysis of actual photo books our partner receives from its customers. With the analyzed photo books and our project partner's long term experience in the field, we identified a first set of semantics derivation rules for Expressive Semantics in photo albums. However, we did not limit ourself to printed photo albums but extended the view to other media types as well. The result is a set of rules that extract common knowledge of Expressive Semantics used in creating multimedia albums. These Expressive Semantics rules are then used to semantically enrich newly created albums.

5.2   Example of Expressive Semantics Rules

For expressing the semantics of a multimedia album, we derived among others the following examples of Expressive Semantics rules: *importance of photos and videos*, expression and annotation of a *semantic concept*, and *page has title*.

Importance of Photos and Videos.

The importance of photos and videos is valuated by comparing their size relative to the album page size. If a photo or video is larger than one third of a page, it is considered more important and a corresponding importance value is increased. Photos or videos smaller than one sixth of a page are valuated less important. The rule can be specified in pseudo-code notation like in the following listing. However, it can also be specified in a more formal way as it has been done in [3] using the Object Constraint Language (OCL).

```
RULE importanceOfPhotoAndVideoRule( medium )
PRE:
  medium.type == PHOTO OR VIDEO
BODY:
  IF medium.size > ( presentation.size  / 3 )
    THEN medium.importance++
  IF medium.size < ( presentation.size / 6 )
    THEN medium.importance--
END RULE
```

Semantic Concept.

Videos and Photos that are arranged on a distinct page of a multimedia album are considered to express a specific semantic concept. Thus, a group is created and the media

assets placed on the album's page are added to the group. This is described by the following rule in pseudo code.

```
RULE semanticGroupRule( page )
  RETURN SemanticGroup group
BODY:
  FOR ALL
    photoOrVideo
  IN
    page.getPhotos() UNIFIED page.getVideos()
  DO
    group.add( photoOrVideo )
  END FOR
  IF pageHasTitleRule( page ) THEN
    group.annotation = page.title
END RULE
```

The rule is similar to the work done by Santini et al. [14, 15] defining semantic groups in the image database El Niño. However, instead that the user is explicitly defining semantic concepts in the database, we derive them implicitly from the authoring process of a multimedia album.

If the page has a title text, one can further derive that this text is a semantic description of the group. The user wants to express that the depicted group of photos and videos constitutes a message conveyed with the title text. Thus, the above rule also checks whether the page has a title by applying the "page has title" rule.

Page has Title.
A title is a text with a minimum font size of 16 points. To express that this text is a title, it is placed in the upper third of the page. The following listing specifies the rule in pseudo code.

```
RULE pageHasTitleRule( page )
  RETURN Boolean hasTitle
BODY:
  IF text EXISTS IN page.getTexts()
  WHERE
    text.position.y < ( page.height / 3 ) AND
    text.fontSize > 16
  DO
    page.title = text
    RETURN TRUE
  ELSE
    RETURN FALSE
END RULE
```

Figure 2 shows an example page of a multimedia album demonstrating the use of the Expressive Semantics rules in xSMART. The page depicts photos and videos from a dolphin watching tour some friends did in Dana Point, Southern California. It shows that the user expresses that the video is a more important media asset in the presentation than the smaller images. By placing the media assets together on one page, they are identified as a semantic concept. This semantic concept is annotated with the text *Dolphin Watching in Dana Point*, as the user expresses this text to be the title of the page by placing it at the top in a sufficiently large font.

We use an easy to understand pseudo-code notation to describe the Expressive Semantics rules in order to reduce the overhead of introducing a model of a multimedia album and describing the rules in OCL. For a more formal definition
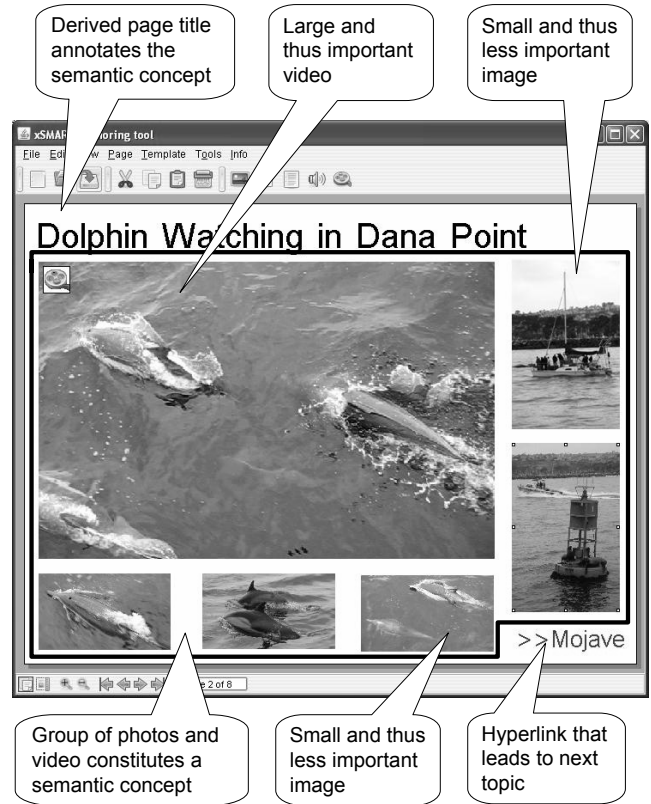

Figure 2: Expressive Semantics rules in xSMART

of a photo album and semantics derivation rules, we refer to our other publication [3].

5.3 Further Rules

Besides the presented Expressive Semantics rules in the previous section, we identified and partially implemented further rules. For example, a rule similar to the *page has title* rule is the *caption of a photo or video* rule. Here, a text that is placed not more than 20 pixels below of an photo or video is considered a caption of the media assets. In addition, the font size of the caption must be smaller than the font size of the (optional) page title. Like the title, the caption semantically describes what is shown by the media asset. Another rule is the *importance of video section*. It is applied when only a part of video is used for a presentation. This part is then considered more important.

A multimedia album can be either a *linear* presentation, which has a fixed, pre-defined temporal course that cannot be changed by the user, or it is a *non-linear*, interactive presentation. The xSMART authoring tool supports the creation of both. By default, the temporal course of a multimedia album is linear. This means that one page of the album is shown after the other once a certain time has passed by (default value is five seconds). Non-linear multimedia albums can be created in xSMART by defining navigational interaction, as shown in Figure 2 with the hyperlink ⟩⟩*Mojave*. A hyperlink can be assigned to any text, photo, or video in the multimedia presentation. By clicking on it, the presentation jumps to a specific page within the album or to an external resource. Thus, a user can express with navi-

gational interaction a narrative structure of the multimedia album that goes beyond the fixed, linear temporal course. Buttons can be added like *Prev* and *Next* jumping to the previous and next page. However, also a button like *Next topic* or *More details about...* can be defined that leads the consumer of the presentation to the next topic in the album or to a section at the end of the album that elaborates the topic in more detail, respectively. The button $\rangle\rangle$*Mojave* in our example leads the consumer to the next topic, a Mojave desert trip the friends did some weeks after.

Defining navigational interaction is a kind of Expressive Semantics as it determines a specific narrative structure of the album. Thus, analyzing it and combining it among others with the semantic concepts identified on the single pages of the album reveals very valuable information. In regard of narrative structures or also called rhetorical structures, there is some very interesting research conducted by Lindley et al. [12]. As navigational interaction also allows for defining hyperlinks to external resources, one can also include the information found on these external web pages for further investigation.

Finally, we see in regard of the rhetorical structures of multimedia albums a high correlation between how to take a video and how to author and structure a multimedia album. Although, they are of different medium type, their narrative structure is very similar. For example, the clips of a wedding video will highly correlate with the chapters of a multimedia album and the used captions and titles.

None of the presented Expressive Semantics rules are hundred percent reliable. Thus, we add a value of reliability to each derived information. This value is depending on the general reliability of the rule as well as the reliability of the input data to that rule. For the reliability of the rules, we have so far only a rough estimation. In future, we would like to base on statistical probabilities of real photo album data.

Although, some of the presented Expressive Semantics rules might seem obvious, it is essential to identify and explicitly define them. Only by this, we get an understanding of what Expressive Semantics for multimedia albums is and can appropriately semantically enrich them. Our goal is to develop more sophisticated rules. These will be backed up with further and more detailed analyses of the photo books received and printed by our project partner. Thus, we would like to conduct many more analyses than we have done so far. We would also like to conduct experiments to see how closely the discovered Expressive Semantics rules pertain to ground truth. As a consequence, this might even lead to modifications and adjustments of the rules discovered so far. For example, it might be discovered that the threshold value of 16 point for title text detection can be represented by a lower value that still covers a very high percentage of correctly determined title texts in real end consumer photo books. Finally, it seems very promising to look into detail how albums for specific occasions like wedding, birthday party, or vacation are created.

5.4 Other Kinds of Semantics

As we described in Section 3.4, Expressive Semantics is closely connected with User Semantics. Thus, in an extension of our system, we would like to integrate User Semantics and take it into consideration when deriving semantics for the authored multimedia album. For example, so far we consider that the story or narrative structure of the multi-

media album is expressed starting with the first page to the last page. However, in other cultural circles, this narrative structure is presented just in the opposite way starting with the last page of the album.

Expressing that some photos (or videos) are more important by drawing them larger than others might lead to a general trend for these photos. The photos that are valuated as important are ranked higher in the selection process of photos to be included in the album. Thus, the photos might be used more often (and again in high resolution) for multimedia albums than others. Considering this use of photos over time shows that their importance more and more reinforces. This is a typical aspect of Emergent Semantics, here connected with Expressive Semantics.

From the use of photos, one can also analyze whether the semantics of a photo is stable or changing. For example, consider a photo that has been used in the context with three other photos on the same page. A new semantic concept is created that contains these photos and is assigned with the title of the page as its semantic description (see "semantic concept" rule in Section 5.2). If the photo is used again in other presentations, another semantic concept will be created and so forth. Analyzing the different concepts in which the photo has been used allows for identifying the emergence of new concepts. This is the case when the used group of photos is constantly changing. If the photo is used in all (or most) pages with a high overlap of the same photos, the semantic concept of the photo may be considered stable.

5.5 Summary

In this section, we showed how Expressive Semantics can be integrated in a system for authoring multimedia albums. We described how these rules are identified and presented the definition of example rules. The rules are used to semantically enrich the created multimedia albums. This makes the previously implicit semantics of the album explicit and allows for a better search and retrieval of the albums. We also presented how Expressive Semantics is connected with other semantics in creating multimedia albums such as User Semantics and Emergent Semantics.

6. CRITICAL DISCUSSION OF SES

Having presented the SES in Section 2, identified the different kinds of semantics and shown their relationships in Section 3, and presented the application of the SES for describing the Expressive Semantics in multimedia albums in Sections 4 and 5, we are now looking at the major challenges involved with the ecosystem. The analysis of the different aspects of our SES, i.e., the investigation of the different semantics and their relationships are far from being finished. Thus, in regard of the research issues of the SES, we identified so far three major challenges that will be discussed below.

From all the different semantics described in the paper, User Semantics is still comparable fuzzy. The User Semantics is in itself a very complex problem as the perception of the Physical World is influenced by a multitude of different factors. User Semantics is not only about, e.g., that a person that was born in South Korea and grown up there has a different understanding of the Physical World as someone who always lived in Canada. However, it is also about the very different facets that exist in human life like that the person might be born in South Korea but than emigrated to

the United States at the age of 12 or was adopted as toddler somewhere in Europe. In order to understand User Semantics better, we need to investigate the single aspects more and need to describe and formulate the facets' dependencies and mutual influences.

Expressive Semantics is a very interesting aspect of the SES as it is related to actions that change the world of semantics, i. e., influence the individual's perception of the Physical World. On another side, it uses basic semantic units to organize them according to some kind of new semantics and hence may have some similarities to Analytical Semantics. It even might find its counterparts in the basic objects and operators we find in Natural Semantics and Analytical Semantics. In any case this is something that requires more thinking.

As presented in the paper, there are clearly different types of Emergent Semantics. Apparently, time is the most important role here. However, as shown there are other factors like context and human experience that play an important role. As a future goal, we like to investigate this further and put together a taxonomy of Emergent Semantics.

7. CONCLUSION

In this paper, we presented an Ecosystem for Semantics. This ecosystem aims at providing a comprehensive approach to holistically understand the complex and challenging problem of multimedia semantics. We have presented the comprehensive theoretical framework, the single semantics involved with it, and their relationships. However, we are still in the early beginnings and still very far from completely understanding the semantics of multimedia. The presented ecosystem provides a good reference to conduct further research. As we have discussed, there are many challenges to be solved and many aspects to think about. Our experience with the SES in authoring multimedia albums are very promising. It shows that the ecosystem is very helpful in order to better analyze, define, and master the complexity of semantics.

Thus, we encourage researchers to take up our Semantics Ecosystem to use it as a reference for describing their own work. In addition, further investigations into particular challenges and aspects of multimedia semantics will be of mutual benefit. The presented Semantics Ecosystem will undergo refinements, specializations, and gain general improvement. On the other hand, researchers benefits from the theoretical framework. This process is natural, it is essential. We strongly believe, that it will help the research community to better understand and master the complexity and challenges of multimedia semantics.

8. REFERENCES

[1] B. Adams, C. Dorai, and S. Venkatesh. Towards automatic extraction of expressive elements from motion pictures: Tempo. In *IEEE International Conference on Multimedia and Expo (II)*, pages 641–644, 2000.

[2] B. Adams, S. Venkatesh, and R. Jain. IMCE: Integrated media creation environment. *TOMCCAP*, 1(3):211–247, 2005.

[3] S. Boll, P. Sandhaus, A. Scherp, and U. Westermann. Semantics, content, and structure of many for the creation of personal photo albums. In *Proc. of the ACM Multimedia Conf., Brave New Topics Session; Augsburg, Germany*. ACM Press, Sept. 2007.

[4] G. Brososky. Signs, symbols and icons (Encyclopedia of educational technology), 2006. Retrieved April 5, 2007, from http://coe.sdsu.edu/eet/Articles/ssilayout/start.htm.

[5] P. Brusilovsky and M. T. Maybury. From adaptive hypermedia to the adaptive Web. *Communications of the ACM*, 45(5):30–33, 2002.

[6] P. R. Cohen and C. R. Beal. Natural semantics for a mobile robot. Technical Report 00-59, Department of Computer Science, University of Massachusettes, Amherst, MA, 2000.

[7] A. K. Dey and G. D. Abowd. Towards a Better Understanding of Context and Context-Awareness. Technical Report GIT-GVU-99-22, Graphics, Visualization and Usability Center and College of Computing, Georgia Institute of Technology, Atlanta, GA, USA, June 1999.

[8] J. Fink, A. Kobsa, and J. Schreck. Personalized hypermedia information through adaptive and adaptable system features: User modeling, privacy and security issues. In A. Mullery, M. Besson, M. Campolargo, R. Gobbi, and R. Reed, editors, *Intelligence in Services and Networks: Technology for Cooperative Competition*, pages 459–467. Springer-Verlag, 1997.

[9] W. I. Grosky, D. V. Sreenath, and F. Fotouhi. Emergent Semantics and the Multimedia Semantic Web. *SIGMOD Rec.*, 31(4):54–58, 2002.

[10] W. I. Grosky, D. V. Sreenath, and F. Fotouhi. Deriving emergent semantics. In *Association for Information Systems, SIG on Semantic Web and IS*, Retrieved January 9, 2007, from http://www.sigsemis.org/articles/grosky/ document_view/, Nov. 2004.

[11] A. Kobsa, J. Koenemann, and W. Pohl. Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships. In *The Knowledge Engineering Review*, volume 16, pages 111–155. Cambridge University Press, 2001.

[12] C. A. Lindley, J. R. Davis, F. Nack, and L. W. Rutledge. The application of rhetorical structure theory to interactive news program generation from digital archives. Technical report, CWI (Centre for Mathematics and Computer Science), Amsterdam, The Netherlands, 2001.

[13] K. Popper. Three worlds [the tanner lecture on human values: Delivered at the university of michigan], Apr. 1978. Retrieved May 7, 2007, from

http://www.tannerlectures.utah.edu/lectures/popper80.pdf.

[14] S. Santini, A. Gupta, and R. Jain. Emergent semantics through interaction in image databases. *IEEE Transactions on Knowledge and Data Engineering*, 13(3):337–351, 2001.

[15] S. Santini and R. Jain. The "El Niño"; Image Database System. In *IEEE Int. Conf. on Multimedia Computing and Systems, Vol. 1*, pages 524–529, 1999.

[16] A. Scherp. *A Component Framework for Personalized Multimedia Applications*. PhD thesis, Carl von Ossietzky University of Oldenburg, School of Computing Science, Business Administration, Economics and Law, Department of Computing Science, 2007.

[17] 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*, pages 802–803. ACM Press, 2005.

[18] A. Scherp and S. Boll. MM4U - a framework for creating personalized multimedia content. In S. Nepal and U. Srinivasan, editors, *Managing Multimedia Semantics*, chapter 11, pages 246–287. Idea Group Publishing, 2005.

[19] A. Scherp and S. Boll. Paving the Last Mile for Multi-Channel Multimedia Presentation Generation. In Y.-P. P. Chen, editor, *Proc. of the 11th Int. Conf. on Multimedia Modeling; Melbourne, Australia*, pages 190–197. IEEE Computer Society Press, Jan. 2005.

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

[21] B. Schilit, N. Adams, and R. Want. Context-Aware Computing Applications. In *Workshop on Mobile Computing Systems and Applications; Santa Cruz, CA, USA*, pages 85–90. IEEE Computer Society Press, 1994.

[22] A. Schmidt, M. Beigl, and H.-W. Gellersen. There is more to context than location. *Computers & Graphics*, 23(6):893–901, 1999.

[23] D. Sreenath, W. Grosky, and F. Fotouhi. Deriving emergent web page semantics. In *Workshop on Implicit Measures of User Interests and Preferences at the 26th Annual Int. ACM SIGIR Conf.; Toronto, Canada*, Aug. 2003.

[24] Stanford Encyclopedia of Philosophy, USA. Karl Popper (Stanford encyclopedia of philosophy), Oct. 2006. Retrieved June 28, 2007, from http://plato.stanford.edu/entries/popper/.

[25] U. Westermann and R. Jain. E - a generic event model for event-centric multimedia data management in eChronicle applications. In *Proc. of the 22nd Int. Conf. on Data Engineering Workshops*, page 106. IEEE Computer Society Press, 2006.

[26] U. Westermann and R. Jain. Toward a common event model for multimedia applications. *IEEE Multimedia*, 14(1):19–29, Jan./Mar. 2007.

[27] Wikipedia. 4 (number), June 2007. Retrieved July 3, 2007, from http://en.wikipedia.org/wiki/4_(number)#In_other_fields.

[28] *3rd Special Workshop on Multimedia Semantics; Pisa, Italy*, June 2005. Retrieved January 30, 2007, from http://www.kettering.edu/∼pstanche/wms05homepage/wms05.html.