Computational models for experiences in the arts, and multimedia

Harini Sridharan Hari Sundaram Thanassis Rikakis

Arts, Media and Engineering Program
Arizona State University
AZ 85281

e-mail: {harini.sridharan, hari.sundaram, thanassis.rikakis}@asu.edu

Abstract

In this paper, we develop formal computational models for three aspects of experiential systems for browsing media – (a) context (b) interactivity through hyper-mediation and (c) context evolution using a memory model. Experiential systems deal with the problem of developing context adaptive mechanisms for knowledge acquisition and insight. Context is modeled as a union of graphs whose nodes represent concepts and where the edges represent the semantic relationships. The system context is the union of the contexts of the user, the environment and the media being accessed. We also develop a novel concept dissimilarity. We then develop algorithms to determine the optimal hyperlink for each media element by determining the relationship between the user context and the media. As the user navigates through the hyper-linked sources, the memory model captures the interaction of the user with the hyper-linked sources and updates the user context. Finally, this results in new hyper-links for the media. Our pilot user studies show excellent results, validating our framework.

Categories and Subject descriptors

H.5.1 [**Multimedia Information Systems**]: *Artificial, augmented, and virtual realities,* H.5.5 [**Hypertext/Hypermedia**]: *Navigation,* H.5.2 [**User Interfaces**]: *Theory and methods, User-centered design,* J.5 [**Arts and Humanities**]: *Fine Arts.*

General Terms

Algorithms, Design, Documentation, Human Factors, Theory, Verification.

Keywords

Art, Experiential models, Context, Semantic nets, WordNet, Hyper-mediation, Implications

1. Introduction

This paper introduces the idea of an experiential document, and we develop computational models for solving one aspect of the experiential document problem. Experiential documents are a subset of the class of problems that deal with experiential systems [25,50]. These systems are concerned with the difficult problem of knowledge acquisition through multimodal interactions. Experiential systems can have a significant impact on multi-modal learning, communication paradigms and content creation tools. We discuss art as a long established mechanism for remediating experiences that can serve as an exemplar for the creation of experiential systems. We discuss the mediation of semantics in art, and the necessity for contemporary forms to incorporate computational models in an organic way.

ETP'03, November 7, 2003, Berkeley, California, USA.

1.1 Related work

The work by Ramesh Jain on experiential computing [25,26] has direct bearing on this effort and we shall discuss his work in terms of the overlap and the important differences. In [25], the author describes experiential computing as follows –

> *In an experiential computing environment, users apply their senses directly, observing event-related data and information of interest … [Experiential environments] don't try to interpret an experience; instead they provide an environment that can be used to naturally understand events…*

The author lists the following as applications – (a) situation monitoring, (b) personalized storytelling, (c) event web and folk computing [26]. He lists several properties of experiential systems – (a) directness (i.e. interfaces are transparent) (b) identical query and presentation interfaces, (c) consider user state and query context and (d) promotion of perceptual analysis. He also presents a simple application that could be easily tailored for the following – data warehousing, activity monitoring and for screening football highlights. While we accept the broad philosophical arguments *for* experiential computing – real-time context-aware, user-centric systems for gaining insight, in [25] there are *no general, principled constructive mechanisms* for creating experiential worlds that provide *specific* insight. We address the issue of mathematical models in this paper.

The informedia group at CMU has also worked on the idea of "experience on demand" [16,53,54] – the main goal was to summarize and visualize data from heterogeneous sources to gain insight. However there are important differences here – to goal was to capture *natural experiences* not create an experiential framework that gains insight. Their representation framework is context and user independent, and does not dynamically evolve with user interaction.

In this paper, we will develop three computational models. First we shall define context using as a union of disconnected graphs, where each node represents a semantic concept, and where the edges define the nature of the semantic relationship. We show procedures to develop two specific contexts – the user and the media. We shall also develop a novel measure of concept dissimilarity. Then we shall show a way to automatically create hyperlinks for *each media* object, using the context of the user and the media annotations. Then we model the evolution of the user context via a novel two-part memory model. The context must evolve as a consequence of the user interaction with the hyperlinked sources. The user studies conducted indicate that the models perform well.

The rest of this paper is organized as follows. Section 2 presents a brief overview of experiential systems. In section 3, we define context and develop models for user and media based context. In section 4, we describe our hyper-mediation strategy. We discuss a two-part memory model in section 5. Section 6 presents two applications of the developed system. We discuss experiments in section 7 and present conclusions in section 8.

2. Experiential systems

We encounter the world we live in, as a multi-modal sensory experience. Let us assume that you are reading this document in your favorite café. As you read these lines with your eyes, you are also paying attention to the environmental sounds — the café, the sounds of the street, at the same time, you can sense the weight of this document on your palms. How can we capture and communicate this encounter electronically? At this moment, there is no way for you to communicate the complex experience to your friends through electronic means – media elements such as images, video clips and a few lines of text will fail to communicate. However, the very same media elements *completely* reconstruct the experience (of the article content, as well the café) for you.

An experience is commonly understood as follows:

> ***ex·pe·ri·ence:*** *the fact or state of having been affected by or gained knowledge through direct observation or participation [1].*

Clearly, this definition is narrow and does *not* adequately encompass the debates on the phenomenology of experience (e.g. as in [29]). The definition however, is useful in creating a computational basis for measuring changes in the knowledge state of the user.

Here are some specific examples of experiences – (a) a child learning to use Lego to build a house, (b) a visit to the beach, (c) the knowledge gained from learning to cook for the first time. Note that each experience involves a gain in knowledge, is multi-sensory, real-time and as well user-centric – it is dependent upon the specific knowledge that each user possess as well as the context of the experience.

In this paper we shall develop computational models that serve as the basis for experiential documents, a subset of the larger class of experiential systems [25]. An experiential document is defined as follows:

> *Experiential documents are real-time, context-aware, user-centric and multi-modal. They cause a variation in the knowledge in the observer, due to direct interaction with the computer mediated environment [25,50].*

This work aims to develop computational models in which these documents can be represented and synthesized. It is important to note that the problem of experiential document synthesis is *not* the same as the problem of re-creating a natural experience. The experiential document, by re-mediating the original knowledge *augments* the knowledge (first-hand or indirect) that exists in the viewer, enabling the viewer to discover connections previously missed. These documents involve the creation of new knowledge frameworks, interactions, visualization / navigation forms, that formally incorporate the constraints due to *existing* communication networks.

Art has been traditionally effective in re-contextualizing semantics into limited media. For example, painters and other artists are very aware of the limitations of the medium that they use [31]. The semantics of that they attempt to convey is informed by the specific medium that they choose, and how it will be mediated.

In the following section we discuss art as a long established system for communicating experiences. We bring forth some of the processes used by art to remediate experiences and we examine the relationship between art and societal state. We thus establish some of the reasons that make art an exemplar for experiential systems. We complete the section by briefly discussing why computational research in experiential systems can allow art to retain an important role in contemporary society.

3. Art and mediated experiences

Immanuel Kant [28] believed that our sense of real experience existed only as a product of our perception – it was impossible for us to stand outside our perception of the world. A human being can not in any way get outside the human experience itself and thus have an outside view of that experience. Thus humans are destined to look at their experiences through the “rose-colored” glasses of their perception mechanisms.

If we are to accept that humans are (consciously / unconsciously) aware of this limitation, then we must expect that human curiosity and the need for evolution will drive them to look for a way to overcome it. They will look for a way to be able to step outside their experience and look at it. Realizing that they cannot do this with the actual experience, they choose the second best alternative – reflect on their experience using a representation. That is one of the major reasons behind the existence of art.

In [12], Braudel points out that art seems to appear only once a civilization has reached a point of maturity that allows it to attempt to organize itself through coherent structures. To do so, a civilization must be able to step outside its current state and contemplate it. For example, when humans began to draw on cave walls they are actually representing (and recording) their experiences and their civilization and are thus able to contemplate them. As the organized hunter becomes part of a civilization, drawings that contemplated that activity began to appear on cave walls. Once humans begin to gather for organized, social, group activities we see drawings of dancing and music making.

The maturity of the symbols used for these representations are also indicative of the maturity of the civilization. When simple sketches of humans begin to be replaced with more lifelike, detailed human figures we can assume that the civilization is looking deeper into its experiences and its state.

As a civilization continues to evolve the experiences that need to be represented grow in complexity and so do the art forms needed for the successful representation of that experience. Once the representation mechanisms reach the maturity of a formal medium, we can begin to talk about mediation of an experience and at the same time we can define the associated art form and style. At this juncture, the feedback loop between art and life also becomes much more complicated and interesting Art is at that point complex enough to be used as a mechanism for deep reflection on life. At the same time contemplating the complexities of art can provide valuable insights into life. We can formulate this loop as a dynamic relationship:

> *The societal impact of the knowledge created by the mediation of an experience through art is directly related to the correlation between the tools and content of the communicated experience and the tools and content of the experiences driving current social evolution.*

3.1 Art and the societal state

Let us look at some examples of art mediating experiences in the context of a civilization using art to contemplate its current state and evolution.

3.1.1 Ancient Greece

In Ancient Greece, at a time of major growth where humans were exploring the power and knowledge that could be gained through the evolution of science, philosophy and the arts the need for exploring hubris was obvious resulting in plays like Oedipus Rex.

An ancient Greek play like Oedipus Rex used forms that allowed the audience to experience the central idea and process of the play (in this case hubris) and thus achieve the final goal of catharsis (cleansing that allows evolution). The audience would partake of the main character's experience through the actor and would come out of the play having experienced the process of hubris.

The origins of tragedy and its evolution show that Ancient Greek society had a very good sense of the intricacies of mediating an experience. In the original Dionysian tragedy, the priest of Dionysus stands in the middle and suffers (in many ways tries to lose his conscience) in order to become one with the God (with the reality that lies outside our conscious experience?). The followers gather around in a circle and try to reach God through partaking of the priest's Dionysian ecstasy. As the early Dionysian tragedy develops during the Athenian civilization into the formalized tragedies of Aeschylus, Sophocles and Euripides, we see the priest replaced by the actor(s) playing the main character(s), the followers by the chorus (that has a number of roles acting as a representation of the people, as representation of common wisdom and sometimes as commentator) and outside all this we actually have the audience. The audience can follow the experience of the main characters and at the same time, through the chorus, a representation or a reflection of its own reaction to that experience. We see that as a civilization develops it organizational intricacies and establishes its formalities its mediation structures undergo a similar evolution. In the case of Ancient Greece, loose, open forms like the early Dionysian tragedy gave way to a highly stylized, multilayered theatrical form.

3.1.2 The renaissance

The Renaissance was an important period when humanity started to re-describe itself. It gradually transformed itself from a helpless victim of circumstance living at a geocentric world to an able observer and analyst of a complex universe. Through the combination of arts and science major breakthroughs were achieved like the development of perspective geometry. Artists like Raphael used geometric perspective to produce paintings in a realistic space allowing the viewers to experience the ability of humans to analyze and represent our world through mathematics and philosophy. The experience of that ability increased the confidence of humans and allowed the western European civilization to start developing its human-centric thought processes.

3.1.3 The modern age

We shall examine a specific painting Picasso's *Guernica*, from the modern era to reveal the relationship between art mediation and the societal state of the time.

Guernica is a strong example of art communicating an experience through an analog medium (painting). The horror and agony of war are directly communicated through the visual composition. Picasso can achieve this communication of experience because he handles perspective, shape, color and light as masterfully as he handles meaning, semantics, memory, context, form and even sociopolitical history. Thus he is able to create a composition where both the medium and the content obey the same processes and serve the same message.

Figure 1: Picasso's *Guernica* illustrates two key ideas – (a) it shows how an artist can skillfully *re-mediate* complex semantics onto media with great limitations and (b) when both the tools (media) and content of a communication are driven by the same goal and with equal mastery then an experience can be successfully communicated and knowledge can be created.

Let us look at some specific examples of this coherent handling of medium, content and message. A big diagonal line starts on the lower right hand corner, from the leg of the woman at the lower right end, cuts through the foreground of the painting (forming an unmovable solid line functioning as the hypotenuse of a orthogonal triangle), and pushes the head of the horse aside. It directs with violence our eyes, through the desperately anxious eyes of the woman, to the brightest point on the painting at the top, a light (or an eye?) that lights the foreground of the painting and could very well also be the explosion of a bomb (but is never explicitly shown as only that). The light/the explosion forces our eyes to run down that same diagonal. At the bottom of this diagonal starts another triangle that engulfs inescapably the body of a person, like an explosion would. Throughout the painting lines coming from the sky pull, strain and dismantle bodies. Mouths open in agony occupy a large part of the foreground. Images of symbolic significance, like the bull, the spear, and the broken sword and hidden images like a concealed human skull in the center of the painting create a second layer of semantic and subliminal messages. Even the printed press, in which Picasso and most of his viewers read about the bombing and saw the photos, is part of the composition.

The structure of the painting makes anyone looking at the painting feel that an inescapable force coming from the sky is tearing apart the human reality we know and expect. There is no need for the viewer to have any personal previous experience of violence in order to feel the agony and brutality communicated by the painting. What make's Picasso's work even more important is that he is communicating the experience of his time using the forms and media of his time. He was thus allowing people of his time to contemplate their experiences, achieve knowledge of their own world and own lives and influence the evolution of their society.

Picasso painted Guernica at a time when paintings were still a major expression and communication avenue for his society. Although electronic machinery and electronic communication avenues like the radio were present, direct physical creation (painting, carpentry, bricklaying) and direct physical communication experiences (like conversations in public forums) or communications through analog means (newspapers, paintings) were still the norm. Furthermore Picasso was living in a society that still functioned through pyramidal structures with messages and orders coming down from leaders. The manipulation of the analog medium of painting to communicate an experience was something that Picasso's audience could directly relate to. The message of humanity's helplessness in the hands of major injustices proliferated from above was a message that resonated with everyone. Hierarchical structures of communication facilitated the dissemination and consequence of a masterpiece created by Picasso. The continuum that exists between Picasso's medium, his message and his time guarantee the success and influence of the work. There are two key points established by this example –

- Consequent art transcodes complex semantics present in the world onto media with great limitations, by understanding the following mechanisms – (a) properties of the media, (b) nature of the mediation and (c) understanding the user and the larger societal context and (d) adopting different forms.
- When both the tools (media) and content of a communication are driven by the same goal and with equal mastery then an experience can be successfully communicated and knowledge can be created. The societal impact of this knowledge is directly related to the correlation between the tools and content of the communicated experience and the tools and content of the experiences driving current social evolution.

3.1.4 Contemporary art and computation

Today, we create through digital means, images, and sounds and we communicate through digital networked media (like cell phones or the internet). These digital media draw their strength and functionality from computational models. However, the expert creators of digital media technologies (the engineers) are not trained in the creation of content. Similarly the creators of content (the artists) are not trained in the creation and exploration of digital technologies and do not have a fundamental understanding of the computational models driving these technologies. It is hard to find people that command the digital medium and its content at the level that Picasso is able to command his analog medium and its content. The computational power of our media is changing our experiences. However that same computational power can only be used to transmit current information or information about experiences of yesteryear.

Our access to information is continuously increasing but when looking for a successful example of communicating an experience through a medium we need to fall back to analog examples like the painting of Picasso. We use the latest digital means to reproduce the image of a painting, collect information about it, and communicate our thoughts to you reading this proposal in possibly different locations and through different representations. In the process we are experiencing many aspects of the digital world but we end up only being able to reflect on an experience from our physical and analog past. Our brain perceives (at least unconsciously) the structural implications of the computational models facilitating this communication but cannot directly connect those structural implications to Piccaso's painting.

The necessity of creating new art forms that are based on the computational mechanisms that drive our everyday experience is clear. As the discontinuum between our media technologies and media content becomes increasingly obvious we are forced to accept that we are a society that excels with information but is unable to reflect on its experiences and thus achieve knowledge of its state. If this ever-growing wealth of information can not be contextualized into experiences, humanity will not have an avenue of true and meaningful interaction with this information. Humanity will remain overwhelmed by the information and frustrated by its inability to function coherently within its own-made environments.

4. The experiential model for multimedia

Research in art as well in the engineering disciplines that deal with multimedia are concerned with the larger abstract question of how representations can re-mediate semantics. Artists have traditionally experimented with the medium until they have created the consistency (i.e. the form) required to re-mediate the semantics relevant to the societal state. Contemporary multimedia research [5,6,18,19,17,40,41,47,48,49,51,56,57,58,59] has focused on creating mechanisms for re-mediating semantics. However they have done so without looking at art creation, the traditional field for experiments in re-mediation.

A framework for communicating experiences is critical for the development of multimedia research, that deals with semantics. As foreseen by McLuhan [32], electronic networks fundamentally alter the relationship of human beings to media. These networks have now largely replaced natural experiences (e.g. a visit to the beach) as our primary sources of information. However, the current generation of multimedia documents use mechanisms for information presentation that are linear, rigid, and reflect forms that are from an old era when such communication networks did not exist – e.g. note the use of storyboards / skims for video summaries (thereby reflecting the ideas behind painted forms / cinema).

Current Multimedia systems present the information in a non-contextual manner (i.e. independent of the user / creator / task context); They also offer limited interactivity. These systems, by isolating the information from the context in which it was created, and by focusing on information reduction, fail to re-contextualize the experience, thus resulting in a tremendous loss of knowledge.

We are in need of models for knowledge spaces that mediate through communication networks and which facilitate new forms of creativity, knowledge exploration and social relationships. These new models will cause us to rethink mechanisms for creating, distributing and storing experiences.

4.1 The communication problem

Communicating experiences is difficult. Let us take a simple example of watching a modern dance performance. If a friend were to ask you via e-mail about your experience, what would you do? You could respond in one of several ways – (a) you could write back saying "it was a wonderful dance performance," (b) send a few photos of the dance, (c) send back a video. However, it is immediately clear that none of these representations will accurately communicate your experience. The communication fails because of the following four reasons:

- *Context:* There is a tremendous loss of context, by the time the message reaches your friend, the recipient. She cannot recontextualize the experience of watching the performance, by simply reading the e-mail, unless she has seen the live performance herself. Importantly, the message is *not* tailored to the recipient. Her notion of modern dance is influenced by her knowledge of the art, and her socio-cultural background.
- *Medium:* The computer mediation (i.e. the act of viewing the message on a computer) is part of the problem! It serves to isolate the recipient from the context of the experience
- *Representation:* The *mechanisms* of representing the experience (image storyboards / video summaries / text / raw signal) are inadequate since they implicitly assume the presence of context (the presence of the creator).
- *Interaction:* There is an implicit assumption here that the acquisition of knowledge is a linear, causal process – i.e. by merely watching the image /video summary, one can absorb "dance."

4.2 The experiential document framework

The ability to communicate an experience using an electronic document is critically predicated on our ability to two things – (a) identify formal elements of such documents and (b) develop mathematical frameworks in which such documents can be rigorously understood and studied.

In our framework, the electronic document is dynamic, context-aware, user-centric and interactive. It changes in real-time due to changes in the user / task / environment context. This shall resemble an highly interactive collage with embedded audio / video / images and text (ref. **Figure 2**) objects. The goal of the document is to propagate specific semantics to then user. These documents *are* experiential – the framework is multi-sensory, context driven, user-centric, interactive and real-time and allows the user to acquire some specific semantics. The experiential document creation also involves the solution of these sub-problems:

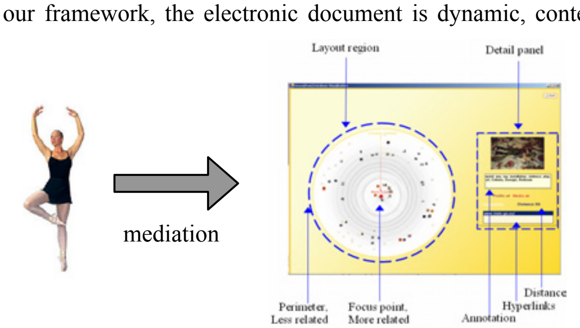
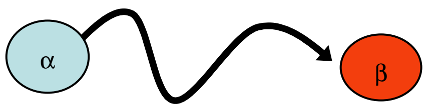


Figure 2: Why is it so hard to communicate the knowledge that we gain from a natural experience using electronic media? What is the framework needed to help re-contextualize the knowledge from a live performance using an multi-modal, interactive electronic document?

- **Context:** We need to develop a context model for experiential documents. This includes determining the formal elements of the context model and the mathematical frameworks needed to represent such elements. The model will need to be adaptive, user-centric. How does each element of the system affect the semantics that we want the user to acquire? What are the limits of the representation? What are the classes of semantics that cannot be acquired for specific experiential representations?
- **Structure:** It is clear from understanding the mechanisms of construction in art [8,11,14,30,31,38,39,52], that the key to *re-mediating an experience* in another media (e.g. dancer captured as a painting), lies in making the viewer re-contextualize the original experience in the new media. Artists often succeed in this re-mediation by making the viewer encounter the *structure* in the original experience. Clearly, we need to develop models for re-mediating structure in experiential documents.
- **Mediation distortion:** Traditional multimedia content summarization / representation mechanisms (storyboards / video summaries / slideshows) make the *implicit* assumption that the semantics to be communicated, exist independent of the user / creator and the mediation. It also assumes that the semantics do *not* change over time. For example consider the problem of trying to communicate the idea of "sounds of the sea." If we just use an image, the semantics due to the sound are lost, and if we just use sound, the chaotic surface property of the sea is lost. The semantics also are dependent upon the scale – consider the difference between communicating this idea using a PDA vs. movie screen. The mediation distortion problem is important to be able to remediate semantics in experiential documents.
- **Synthesis:** The goal of experiential synthesis to automatically create an interactive context driven user-centric, multi-modal document. The document should allow the user to absorb the specific semantics that she needs. For example, assume that the user wants to understand a particular dance style, using multimedia elements on a computer. The experiential synthesis problem is to determine right multi-modal representations to allow *this* user to acquire this multi-modal knowledge given context, content, user profile and method of user interaction. We believe that this can be formulated as a non-trivial optimization problem with constraints. The constraints include those for context (user/ task / mediation, content semantics), structure (in the data to be re-mediated) and time. Another way to visualize the synthesis problem is to consider the user to be at a certain knowledge state $K(\alpha)$ and wishes be a different knowledge state $K(\beta)$. Then, the optimization problem involves determining the optimal experiential path (i.e. visualization, interaction) that satisfies the knowledge constraint while minimizing a resource e.g. time.

$$
\begin{aligned}
&E = \{ p \mid \Delta k(p) \geq K(\beta) - K(\alpha) \}, \\
&p^* = \arg\min_{p \in E} O(p),
\end{aligned}
$$
<1>

where, K(α) and K(β) represent the knowledge states of the user E is the set of valid experiential paths that satisfy the knowledge constraint, $\Delta k(p)$ is the knowledge gain due to a specific experiential path p. O represents an optimality objective function that could for example measure the time taken to gain the knowledge required by the user. This is illustrated in **Figure 3**.



$$\Delta k(p) \geq K(\beta) - K(\alpha)$$

Figure 3: An experiential system determines the optimal context-adaptive path p that leads to an *variation* in knowledge in the observer, while minimizing an objective unction.

Artists excel at communicating experiences [7,30,39,52] with limited media. Through interaction with many new media artists at Arts Media and Engineering program at ASU, it became clear that experiential documents in general, must be multi-modal, context-aware, dynamic, highly interactive and critically, *real-time*. Over the next few sections, we present in detail the three components that comprise our experiential document model: (a) context models (b) serendipity and (c) memory models for context evolution. Note that these models can evolve to deal with the larger class of experiential systems for performances (ref. Section 3.1.4).

5. Context

In this section we : (a) define context and come up with a formal model of context. We will show the use of WordNet to be integral to our model. (b) determine a procedure for finding concept dissimilarity, (c) and finally, we shall apply our context model to develop models for the user and the data.

The context models developed in this paper will be tested on a database of media objects (images / videos / sound / text) that have been annotated.

5.1 What is context?

The Merriam-Webster dictionary defines context as follows:

> ***con·text:*** *the interrelated conditions in which something exists or occurs.*

These conditions could be information got from the media being heard or viewed, facts related to the environment (e.g. “Delhi is the capital of India”, personal information about the user (e.g. “female”, “jazz” etc.), the user’s past actions, the time, the place etc. In our framework, we define context as the various conditions that interrelate to impart meaning to the media objects accessed by the user. For example, the text (which is a form of media) ‘Solaris’, ‘Ice-Age’ and ‘Speed’ could refer to any of their traditional linguistic semantic connotations. However, when certain conditions like the environment being a movie theatre combine to establish the context of films, we can fixate the semantic notion associated with these words to be film titles.

There has been much work using the idea of context primarily in the area of context-aware / ubiquitous computing [24,22,23,44]. The work there focuses on *application use* and hence is primarily concerned with contextual information such as location, identity,

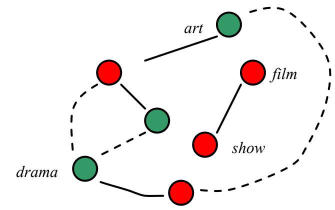


Figure 4: An semantic net defining the user’s interests. In general, the user context will be a union of such networks

activity and time [22]. However, the notion of context in multimedia is concerned with *semantic* inter-relationships between concepts, which can be arbitrary. Secondly the relationships that wish to explore in multi-media are linguistic, statistical as well as common sense rules such as from the CyC database [4]. There has been some work in multimedia analysis / wearable computing [20,21,34,43] – however, there is no *formal* definition of context, specific properties are used for pattern detection.

In order to come up with a formal model of context we introduce the notion of a semantic-net.

5.2 Semantic-nets

A semantic-net is a graph $G = \langle V,E,W \rangle$ where the node $v_i \in V$ represent the concepts. A concept node is associated with a specific instance that could be an image, video, an audio segment, text etc. The edges $e_{ij} \in E$ of the graph represent the semantic interconnection between any two concepts. The strength of the relationship between the nodes i and j, is given by $w_{ij} \in$ W. For example, the words “film” and “show” have a “is-a” linguistic relationship. More generally, we allow semantic-nets to consist of disconnected sub-graphs. In **Figure 4**, we show an semantic net that defines a user’s interests.

5.3 A formal model for context

We define the context be the union of semantic-nets:

$$C = \bigcup_{i=1}^{k} G_i \qquad \text{<2>}$$

where C is the context, k is the total number of semantic-nets, G_i is the i^{th} semantic-net with the semantics of the nodes and vertices defined in the previous section. For example, the user context could be defined as the union of semantic-nets, each of which specifies interests, occupation, ethnicity etc.

The overall *system* context is defined by the inter-relationships amongst the various aspects of the system – the user, the environment, the allowable interactions etc. Each of these aspects are modeled as a specific context (ref. equation <2>) Thus, the system context is given by

$$C_s = \bigcup_{j}^{N} C_j \qquad \text{<3>}$$

where, C_s is the overall system context, C_j is the j^{th} context (e.g. the user context) in the system, and where N is the number of specific contexts. We will follow this interpretation of context through the rest of this paper.

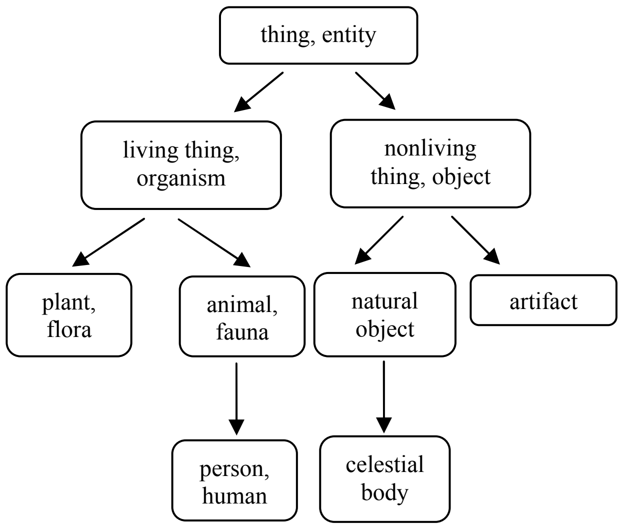


Figure 5: A hierarchy in the WordNet ontology

In our framework, the concepts are represented as text, as opposed to the more general framework in [9], that deals with multiple media for concept representation. We have focused on text because of two reasons – (a) we are working with text annotated media and (b) there are textual ontologies such as WordNet [33] created by linguists which specify semantic relationships amongst concepts. Note that no such universal agreement exists for non-textual media. We shall discuss the WordNet ontology in the next section.

5.4 WordNet

WordNet is a differential lexicographic arrangement of words created by linguists which specifies semantic relationship between the concepts as represented by words. The semantics associated with each word is represented by a *synset* or a synonym set that are a group of words which sufficiently characterize the semantics associated with the word. For example, the idea of *film* as in motion-picture is described by the synset: "movie, film, picture, motion picture, picture show" which is sufficient to differentiate it from the synset: "film, photographic film" which is a different sense of the same word.

These sysnsets are the nodes of an ontology where the edges represent relationships amongst the synsets. WordNet relationships can be synonyms/antonyms, hypernyms/hyponyms (A word w_1 is a hyponym of another word w_2 (the hypernym) if the relation is-a-type-of holds between the meaning of w_1 and w_2 – for example, apple is a fruit) or meronyms/holonyms (A word w_1 is a meronym of another word w_2 (the holonym) if the relation is-part-of holds between the meaning of w_1 and w_2 - for example, wheel is a part of a car). In our framework, we are interested in the hypernym/hyponym relationship which creates a hierarchical semantic organization between concepts. Figure 5 illustrates an example of such a hierarchy.

Integrating this in our framework, we define the nodes belonging to semantic-nets using synsets, the relationships between the nodes being given by the generalization/specialization (i.e. the hypernym/hyponym relations in WordNet). The edge weights are specified using concept distances that are derived using concept implications as discussed below.

5.5 Concept Dissimilarity

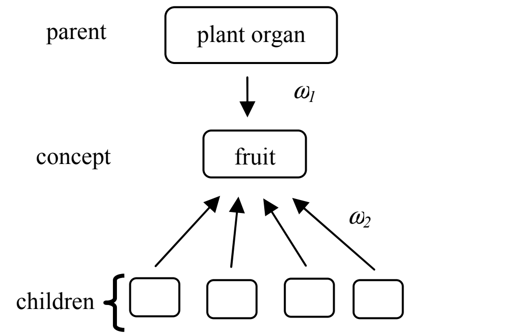


Figure 6: A concept is implied by its parent and its children. The implications of the children are considered independently of the parent.

In this section, we shall determine a procedure to compute the dissimilarity measure between any two concepts in the semantic-net. The measure is *not* a metric, since the semantic-net has a specific generalization/specialization relationship associated with it. This causes the symmetry requirement of a metric not to hold. We shall compute dissimilarity using an intermediate notion of implication.

A concept α (e.g. fruit) has two entities associated with it – the parent concept, and the children concepts (ref. Figure 6). Each entity implies the concept with a different weight – ω_1 (the parent) and ω_2 (all the children). Hence the *implication* that the concept α is true given that another concept β is true is computed as follows:

$$\begin{aligned} I(\alpha \mid \beta = T) &= \omega_1 I(parent \mid \beta = T) + \frac{\omega_2}{k}\sum_{i=1}^{k} I(c_i \mid \beta = T), \\ I(\beta \mid \beta = T) &= 1, \\ \omega_1 + \omega_2 &= 1, \end{aligned} \quad \text{<4>}$$

where I is the implication strength, and k is the number of children of the concept α, c_i is the i^{th} child of the concept, and where ω_1 and ω_2 are the weights attached to the implications of the parents and the children respectively. Note the following:

1. Implications of the parent and the children nodes are evaluated independently of the concept α, to prevent cycles.
2. Implications of the root node, or leaf node with respect to concept β can only be equal 1 in the case of identity:

$$I(\alpha \mid \beta = T) = \begin{cases} 1 & \alpha = \beta \\ 0 & \alpha \neq \beta \end{cases} \quad \text{<5>}$$

   where α is a either a root node or a leaf node in WordNet.
3. The weight ω_1 is computed to be inversely proportional to the number of children of the *parent* concept, α. i.e.

$$\omega_1 \propto 1/m \quad \text{<6>}$$

   where m is the number of children of the parent.
4. The distance between the two concepts is now easily determined as follows:

$$d_U = 1 - I(\alpha \mid \beta = T)$$
$$d(\alpha \mid \beta = T) = d_U / \sqrt{f_\alpha \bullet f_\beta}$$ <7>

where d_U is the un-weighted distance between the two concepts. f_α and f_β, represent normalized knowledge priors for concepts α and β. The priors are used to re-weight the distance. These could be set by the user, as part of her context model. If these values are unknown we just use $f_\alpha = f_\beta = 1.0$. Note that these values could also be determined using the normalized frequency of occurance of the concept.

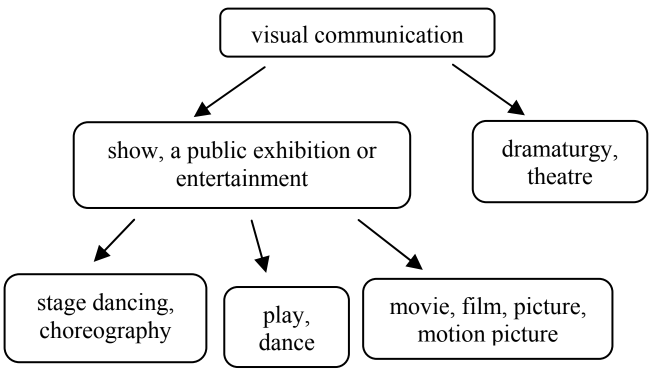


Figure 7: Using the WordNet hierarchy to find the dominant concepts in the user profile.

Again, the measure is anti-symmetric, as expected. This recursive formulation is (ref. equation <4>) is computationally expensive, so in practice we approximate the implication value using a pruned tree [46].

5.6 Context models

Now, we discuss the methodology for modeling a specific user and the media.

5.6.1 Modeling the user

In this section, we show how we construct the user context. Modeling the user context is important as it enables us to create an unique, user-centric experience. We adopt a semi-automated procedure discussed below, to create the user context.

1. The user inputs to the system, details about herself – gender, ethnicity, age, profession, cultural interests etc.. Each of these entities is input as text. Each textual entity input is generalized using the WordNet hierarchy.
2. The entities that fall under one generalized synset are grouped as belonging to one concept. This concept is said to form a cover for all the entities that it generalizes. For example, if the user profile says that she is an artist and a dancer and is interested in theatre, drama, art and cinema, then the concept of visual communication becomes dominant in his or her profile according to the WordNet hierarchy as illustrated in Figure 7.
3. A semantic net is created for each aspect of the user input – ethnicity, gender, interests etc. using the textual entities The user's context is then the union of such semantic nets (ref equation <2>).

The media accessed by the user needs to be modeled, as this also part of the system context. We discuss this in the next section.

5.6.2 Modeling the media

The media elements in the database can be arbitrary (audio, video, text or image) and we assume that they are annotated with text. For audio and video clips and images, we create semantic nets by generalizing the annotations using WordNet. The following steps detail how we build up the semantic-net for the text (text documents or annotations) in the database.

The following steps detail how we build up the semantic-net using the text documents in the database.

1. Each non stop-word (words that are so commonly used that they no impact on the import of a statement; for example, the words 'is', 'the', 'of' etc. are stop words) that is a noun in the text becomes an elementary entity. We consider only nouns in our framework as the annotations of a media are mostly nouns.
2. Then we generalize each elementary concept and determine the dominant concept cover. The list of covers is pruned by removing concepts that cover fewer than m entities. Thus:

   $$G = \{v_1, v_2, \ldots, v_k\}, \text{ where } |v_k| \geq m$$ <8>

   where, G is the semantic-net, v_k is a dominant concept belonging to the semantic net and m is a constant and where $|v_k|$ is the the number of elementary entities covered by v_k.
3. An elementary entity itself is made into dominant concept if

   $$f_e / f_c > \alpha$$ <9>

where f_e is the frequency of occurrence of the elementary entity, f_c is the sum of the frequencies of all the elementary entities in the parent concept, and α is a constant less than 1. In other words, an elementary entity is made dominant if its frequency of occurrence is a fraction α of the sum of the frequencies of all the elementary entities in the concept to which it belongs. The values of m and α have been optimized to 3 and 0.75 respectively by experiments. The union of all these semantic-nets forms the media context.

6. Serendipity

We rarely acquire knowledge, particularly through electronic documents in simple linear way. For example, if you read this document as an electronic file on your computer, it is very likely that you did some or all of the following – briefly checked e-mail, browsed some web pages and perhaps made yourself some coffee. Each of these *serendipitous* activities led to an increase in your *overall* knowledge. Serendipity has also been explored in other contexts [13,37].

In recent work we decided to introduce the idea of serendipitous acquisition of knowledge as part of the solution for browsing the Ralph Lemon dataset (Section 9.1) using context-aware hyper-mediation [46]. Hyper-mediation does two things – (a) It allows for a user-context constrained exploration of information related to media and (b) it introduces serendipity into the user's exploration of the media. Hyper-mediation is *one* approach to creating a serendipitous environment – clearly other mechanisms exist as well.

In our approach, we use Google [2] to automatically generate hyper-links. This is done by taking into account the user profile, the semantics of the media and the semantic relationship between the media item and the user profile. We use the first hyper-link returned by Google.

For each non stop-word *w* that is part of the annotation of a media object (images, video, audio, text), we proceed as follows:

1. If *w* is present in the user context as an elementary entity, then *w* along with the dominant concept cover is sent to Google as a query. The first hyperlink is chosen and is added to the list of hyperlinks associated with that media object.
2. If *w* is absent as an entity in the user profile, then the dominant concept closest to it is found using the concept dissimilarity (ref. equation <7>). Then *w* and the dominant concept is used to query Google, and the first hyper-link is again associated with media object.
3. If the concept distance is very large from all the concepts in the user context, then we adopt the following strategy:
    a. If the word is a common noun, it will not be hyper-linked at all.
    b. If it is a proper noun, the word alone is used to query Google. Proper nouns are hyperlinked even if they do not form a part of the user semantic net because it is natural that the user might be interested and would like to know more about these words.

For example, let us assume that the word "Solaris" appears both as a media annotation and as an elementary entity of the "film" concept in the user profile. Then, the media is hyper-linked to the first link returned by Google for the query "Solaris + film." The hyperlinks are thus relevant to the current context. Note that the word 'Solaris' alone as the query would return the home page of the Sun Microsystems OS - Solaris which is totally out of context.

At the end of this procedure, we have a hyper-link for each non stop-word annotation of the media in the database. As the user clicks on the generated hyperlinks, she is taken to a page in the World Wide Web and she spends some time on this page. Now that the user has absorbed information from that page, this newly acquired information too should become a part of the user context. In the next section we discuss a model for user-context evolution.

7. Context evolution

In this section we discuss how user context evolves over time in a way that is analogous to the human memory. Memory models are important as memory forms an essential part of communicating an experience by the way of associations.

7.1 The leaky bucket model

We now explain a computational model of memory and how it incorporates concepts acquired by the user navigating the hyperlinks, into the user context.

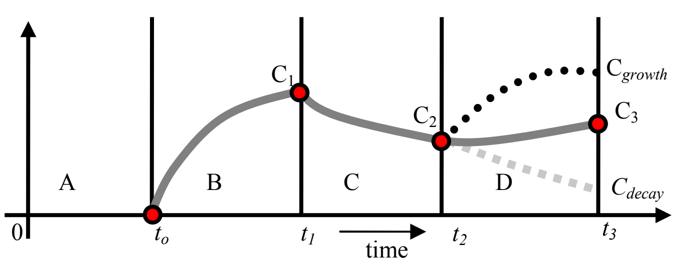


Figure 8: The life of a concept though four stages.

This is the intuition behind our memory model. For each hyper-linked page visited by the user, the experiential system creates a dominant semantic net from the text; it also measures the time spent by the user. Once the user comes back to exploring the media, it is natural that certain new concepts be introduced in her user context (newly gained knowledge), certain concepts get reinforced (due to associations with similar concepts in the user's context), and certain other concepts decay (put behind over time).

In our framework, we develop a two part model for memory: *permanent* memory and a *transient* memory store. The permanent memory contains all the entities that the user specifies at the beginning of her interaction with the system. These permanent concepts are assumed not to change with time. The transient memory contains all the concepts that were automatically extracted by the system from all the pages visited by the user. The transient memory model is implemented using the *leaky bucket model* of memory – concepts automatically acquired by the system are slowly lost over time at a fixed rate, unless they are reinforced by the user visiting related pages. This concept reinforcement is analogous to forming associations of a current event with an earlier event that generated feelings of a similar experience. This idea is best explained with an example.

We shall examine the growth and decay of a concept *K* which is absent from the initial user context. We will analyze four stages in the life of a concept, *K* as illustrated in **Figure 8**. that demonstrates its life and growth as per the transient memory model. We explain progression of the weight of the concept in the user memory model, in each of four stages below:

A : The concept *K* has not been encountered as yet and so its weight is zero.

B : At time t_0, the user visits a web page that emphasizes concept *K* with a normalized frequency f_1. The viewer views the page till time t_1. The concept weight grows[1] to C_1 according to the growth function given as follows:

$$\begin{aligned} C(t) &= 1-\exp\left(-\beta f_1(t-t_0)\right), \\ C_1 &= C(t_1), \end{aligned} \qquad \text{<10>}$$

C : At time t_1, the user moves over to another page (on which he stays till time t_2) which does not contain the concept *K* in it. Hence the concept decays through that period and is given by

$$\begin{aligned} C(t) &= C_1\exp\left(-\alpha(t-t_1)\right) \\ C_2 &= C(t_2) \end{aligned} \qquad \text{<11>}$$

α and β are constants (β being greater than α as the knowledge imbibing rate should be greater than knowledge decay rate.

D : At time t_2, the user again encounters a page with the concept *K* in it emphasized by a normalized frequency f_2. The resulting function is the sum of the growth function (stage B) and the decay function (stage C):

$$\begin{aligned} C_g(t) &= C_2+1-\exp\left(-\beta(f_2+C_2)(t-t_2)\right), \\ C_d(t) &= C_2\exp\left(-\alpha(t-t_2)\right), \\ C(t) &= \left(C_g(t)+C_d(t)\right)/2, \\ C_3 &= C(t_3), \end{aligned} \qquad \text{<12>}$$

[1] Note that there is an assumption that the user is paying equal attention to all the concepts in the page.

where, C_g and C_d represent the growth and decay function. Note that the growth function is dependent on the old weight C_2 and the weight of the concept seen – f_2. Hence the concept weight at the end of stage D is the average of the growth and decay functions.

In **Figure 8**, we examine the four stages of a concept (ϕ) that grows automatically, as a result of user interaction – (A) ϕ absent initially, (B) ϕ grows after being encountered for the first time, (C) user leaves page – ϕ decays and (D) ϕ is encountered again – there is growth (round dots) as well as decay (square dots), and resulting in reduced growth. The hyper-mediation of *all* media objects needs to be revised as a consequence of the new concepts being added to the user's transient store. This is because the distance of the annotations of the media with respect of new user context has changed. Then, we again use the strategy outlined in section 6 to regenerate the hyperlinks.

To summarize – we have shown how to create formal models of user and media context, and a context constrained hyper-mediation strategy. Finally, we have shown a leaky memory model to incorporate changes in the context that take place due to the user interactions.

8. Experimental validation

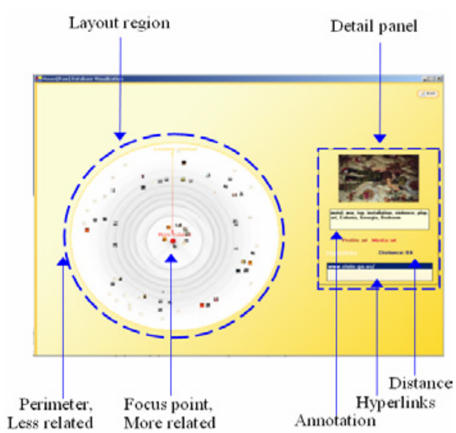


Figure 9: The visualization interface used in our experiential system.

Evaluating models for experiential systems is challenging. We conducted a two-part evaluation of our system : (a) evaluating the quality of the hyper-links and (b) evaluating the change in the hyper-links as the user context changes. An assessment of the quality of the hyper-links would indirectly be an assessment of (i) the user and media context model developed (ii) the query formulation for hyper-mediation and (iii) the concept distance function used. Evaluation of the change in the hyper-links is an implicit evaluation of our memory model and concept distance.

Table 1: Average of test scores from five users evaluating the quality of the hyperlinks

Questions	Score
Hyperlink quality [1-7]?	4.96
Changed hyperlinks are more interesting [0-1]?	0.61

The media (images / video/ audio clips and text) used in this experiment was taken from the house[raw] project [3]. Each media element was annotated by the creators of the project. We used a user-interface that displays all media in a collection with their positioning based on their concept dissimilarity from the user profile (ref. **Figure 9**). Images are placed on concentric circles with the images nearer to the user profile closer to the centre. The user can choose images with a mouse and the details (hyper-links, distance to user profile, concept relationship) are then displayed in a side panel. The user can follow the hyperlinks within the system using an embedded browser, and subsequently return to exploring the media data.

We evaluated the models through a pilot user study with five graduate students. We conducted two experiments. We first evaluated the quality of the hyperlinks. The users were made to choose any five images that interested them from the visualization and they were asked to rate the hyper-link quality (scale 1 – 7). Then, they were also asked to determine misses (annotations not linked) and false alarms (linked but not relevant) annotations. Note that the hyperlinks are unique to each user (ref. Section 6).

The second experiment evaluated the change in the hyperlinks after the user had followed the original hyperlinks and browsed the web. After browsing for a fixed period, the user was asked to get back to the same to the same set of images as in experiment one. Then, the user was asked (yes/no) if the changed hyperlinks were more interesting.

Table 2: Accuracy of semantic nets and hyper-mediation

Accuracy	Score
Misses	12.2%
False alarms	4.0%

The user study results shown in **Table 1** and **Table 2** are very encouraging. Both tables show the results averaged over five users. The hyperlink quality (4.96) was evaluated using the Student's t-test, and could occur by chance only with probability $p < 0.15$. The users found most of the changes useful. The very few misses and false alarms are very good indicators of our semantic-net framework for context modeling. There were two sources of error – (a) most of the misses occurred because the user profile was not exhaustive enough and (b) there were spelling errors in the annotations, leading to incorrect hyper-links. The first error could be corrected by incorporating relevance feedback while the second error can be easily fixed using a spell-checker. A few errors were also due to the fact that only nouns in the annotations were taken into consideration. The simple solution lies in allowing other word forms also.

For a specific user, it is quite possible that for many media, hyperlinks are not created due to the media annotation being irrelevant to the user profile. A solution to this is that at such instances, we could hyperlink the annotation (by querying Google on the words in the annotation alone) so as to provide means for context-independent navigation.

9. Applications

We have applied our experiential models in two scenarios – (a) context adaptive visualization scheme (b) an experiential annotation system. We now briefly describe both projects below.

9.1 Context adaptive visualization

The context adaptive visualization scheme uses semantically rich dataset – from the Ralph Lemon project that deals with the African American experience. We shall first describe the project mandate and a then provide a brief overview of the project. More details can be found in [42].

9.1.1 The Ralph Lemon project

(Collaboration with Ralph Lemon and John Mitchell, Dept. of Dance). Ralph Lemon is a famous modern dancer, who is working on the third part of his *Geography* dance trilogy named *House.* The dance deals with the idea of representing African-American identity through movement. Ralph Lemon toured the southern states in the U.S. and collected multimedia material. The multimedia research mandate for this digital content:

- How to re-contextualize the African-American experience (i.e. the knowledge / understanding of "African-American-ness") for the user through a real-time, interaction with the media? Additionally, the experience that is created must be user-centric.
- The semantics of "African-American-ness" have changed over time – how to create an interaction that captures the temporal evolution of semantics?

The project mandate dictates that he user interact with the data in a context-aware user centric manner – this was the direct motivation for the context aware adaptation scheme.

9.1.2 Context adaptation

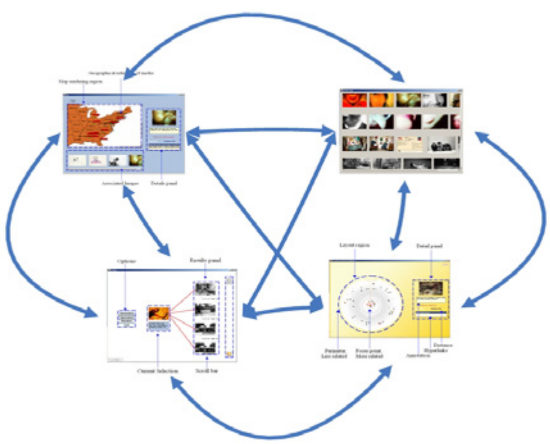


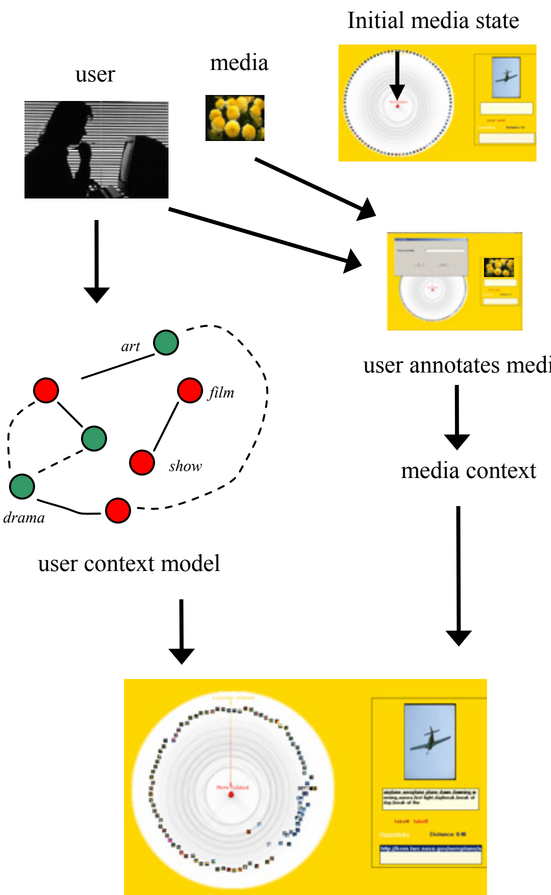
Figure 10: User's preferences and the semantics of the data being viewed determine the transitions in the visualization schemes

The project [42] deals with adapting the visualization scheme for viewing the dataset of an African American dancer depending on its utility. The dataset consists of media files collected from various locations of the civil rights movement and the project is a step towards the goal of conveying to the user, the experience of being an African American during the civil rights movement.

The utility of the visualization scheme chosen is derived from the media annotations and the user profile. For example, if a user is browsing through media objects that are all associated with places of interest and the user is an extensive traveler, she might be interested in a visualization scheme that uses maps.

The procedure employed is briefed as follows: A set of visualizations are developed, each of which are applicable in a different scenario and for a different set of user tasks. Thus, each visualization has specific semantics associated with it. Using the concept dissimilarity measure explained in section 5.5, the distance between each visualization scheme and the user context (built from the user profile) is found. This distance measure attributes a weight to the visualization which is a measure of how close the visualization is to the user profile.

A specific visualization is picked to view the dataset, that depends upon the weight associated with each visualization and the semantics of the data being viewed (e.g. location is concerned with geographical concepts). Each of the visualizations also incorporates the hyper-mediation scheme explained in section 6 so that as the user navigates through the hyper-linked sources, the



The new media state, where each media now is hype-linked to sites related to the media context as well as the user context. The user can now provide relevance feedback and the media context gets updated anew.

Figure 11: The experiential annotation system. The goal of this system is to create an real-time annotation environment that *motivates* consumers to annotate, by providing them insight about their relationship to the media semantics.

memory model captures the interaction of the user with the hyper-linked sources and updates the user context. This in turn will affect the visualization being chosen the next time the user browses through the media objects. The user study on the application showed that certain visualizations do aid in the completion of certain tasks much faster than others. The users were also more interested in an adaptive visualization technique rather than in a static visualization. This also goes to prove that different visualizations for different data-set and tasks will enhance the users experience with the data.

9.2 Experiential annotation

Annotation is a key problem in multimedia [15,55]. However, while prior work has extensively focused on better computational models for annotation, there is very little incentive for ordinary consumers to annotate the media. This is because the *result of the annotation is a traditional visual layout scheme – there is little return on the enormous amount of time spent by the consumer on annotation.*

We have attempted to address this issue in our project on experiential annotation [45]. The project deals with creating a new user centric annotation technique that aims at making the process of annotation more interesting to the user. The user, in the process of annotation will gain fresh insight into the data, thus motivating her to participate in the annotation process.

The annotation system is semiautomatic – the user annotates a subset of the media/images in a database and these annotations are propagated to the rest of the database based on the structure (low level features of the image) and semantics (using the WordNet ontology) of the images. The flow diagram (ref. **Figure 11**) illustrates the process. After the user annotates an image, she fixates on the sense of the words in the annotation. The distance of each image from the annotated image is calculated as a function of the low-level feature distance and the WordNet distance (ref. concept dissimilarity in section 5.5) between the two images. A threshold is fixed for this distance and the annotation is propagated to all images whose distance from the manually annotated image falls below this threshold and which are close to the user profile.

In the annotation system user interface, all these images are shown in relation to the semantic distance with respect to the user profile. The annotations are also hyperlinked based on their relationship to the user profile. As the user begins to annotate the media, the system begins to start generating hyperlinks for the media. The user interaction with these hyperlinks changes the user context, thus altering the distance of each media element from the center (represented by the user profile). The visualization, and the hyper-links thus change dynamically (also ref. section 6).

In a pilot user study with ten graduate students, the users found the new annotation mechanism more useful than the traditional annotation interface to be found in modern operating systems.

10. Future work

We would like to examine limitations of our approach towards modeling context, that we would like to address in our future work:

- *Text only:* The present formulation of only using text needs to be expanded and we plan on improving on the work on MediaNet [10,9], to include non-textual (e.g. audio / visual) concepts. While we only use linguistic relationships at the moment, we are also planning on using probabilistic models in the form of Bayesian belief network and other graphical models [27,34,35,36]. Finally, we shall formally incorporate common-sense knowledge using OpenCyc [4] in the framework.
- *Explicit representation:* Experiences are often pre-linguistic – for example the experience of going to a beach, cannot be adequately captured by an explicit semantic net. We would like to like to focus on developing *implicit* context models that can represent a continuous variation of semantics. Indeed semioticians realize that many forms of cultural communication cannot have a discrete linguistic basis [31].
- *Expressive power*: Finally, the models that we have developed do not yet have metrics that enable us to quantitatively measure expressive power of the framework. For example, a bound (or a covering theorem) on the forms of knowledge that can be expressed using such context models would be helpful.

11. Conclusions

In this paper, we have presented computational models for three formal components of an experiential document. Experiential documents are a subset of the larger class of experiential systems (including multimedia art performances). These systems deal with the acquisition of knowledge via a direct, real-time interaction with the environment. We also discussed the relation between mediation of semantics in the arts and the societal state, and postulated the need for art mediation being based on computational mechanisms.

We presented a specific communication example, and examined how problems due to context, mediation, representation and interaction cause a loss of communication. Then, we presented an experiential document framework that comprises four sub-problems – (a) context models, (b) structure discovery, (c) mediation distortion and (d) experiential synthesis. In this work, we have only focused on context models.

We showed how context can be defined as a union of semantic-nets, and then we showed how user and media context can be constructed. We also derived a novel concept dissimilarity measure. Then we discussed the importance of hyper-mediation and showed a procedure to create hyper-links using the media annotation and the user context. The hyper-links were automatically generated by Google. Finally, we derived a two-part context evolution model analogous to human memory. The evolution model updates the user-context, thus resulting in new hyper-links for the media in the database. The user study indicates that the system performs well – thus implying that the models could serve as the basis of experiential systems.

12. Acknowledgment

The authors wish to thank John Mitchell, Ralph Lemon, Siew Wong, Jaikannan Ramamoorthy for their collaboration on different aspects of this work.

13. References

[1] *Merriam Webster Dictionary* http://www.m-w.com.

[2] *Google* http://www.google.com.

[3] *house[raw]* http://isa.asu.edu/projects_house.html.

[4] *OpenCyc* http://www.opencyc.org.

[5] L. Agnihotri, K. Devara, T. McGee, et al. (2001). *Summarization of Video Programs Based on Closed*

Captioning, SPIE Conf. on Storage and Retrieval in Media Databases, San Jose, CA, 599-607, January 2001.

[6] B. ARONS (1994). *Pitch-Based Emphasis Detection For Segmenting Speech Recordings*, Proc. ICSLP, Yokohama Japan, 1931-1934, Sep. 1994.

[7] R. ASCOT (1966). *Behaviourist Art and Cybernetic Vision*. Cybernetica **9**: pp. 247-264.

[8] R. BARTHES and S. HEATH (1988). Image, music, text. New York, Noonday Press.

[9] A. B. BENITEZ, J. R. SMITH and S.-F. CHANG (2000). *MediaNet: A Multimedia Information Network for Knowledge Representation*, Proceedings of the 2000 SPIE Conference on Internet Multimedia Management Systems (IS&T/SPIE-2000), Boston MA, Nov 6-8, 2000.

[10] A. B. BENITEZ and S.-F. CHANG (2002). *Multimedia Knowledge Integration, Summarization and Evaluation*, Proc. of the 2002 International Workshop On Multimedia Data Mining in conjuction with the International Conference on Knowledge Discovery and Data Mining (MDM/KDD-2002), Edmonton, Alberta, Canada, July 2002.

[11] J. D. BOLTER and R. GRUSIN (1999). Remediation : understanding new media. Cambridge, Mass., MIT Press.

[12] F. BRAUDEL, R. D. AYALA, P. BRAUDEL, et al. (1998). Les mémoires de la Méditerranée : préhistoire et antiquité. Paris, Editions de Fallois.

[13] V. BUSH (1945). *As We May Think*. The Atlantic Monthly. **176:** 101-108, http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm.

[14] J. CAGE (1966). *Diary: Audience 1966*. A year from Monday, Middletown CT, Weslyan University Press, 1966.

[15] E. CHANG, K. GOH, G. SYCHAY, et al. (2003). *CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines*. IEEE Transactions on Circuits and Systems for Video Technology **13**(1): 26-38.

[16] M. CHRISTEL, S. STEVENS, T. KANADE, et al. (1996). *Techniques for the Creation and Exploration of Digital Video Libraries*. Multimedia Tools and Applications. B. FURHT. Boston MA, Kluwer Academic Publishers.

[17] M. G. CHRISTEL, M. A. SMITH, C. R. TAYLOR, et al. (1998). *Evolving video skims into useful multimedia abstractions*, Proceedings of the SIGCHI conference on Human factors in computing systems, Los Angeles, California, United States, 171-178, 1998.

[18] M. G. CHRISTEL (1999). *Visual digests for news video libraries*, Proceedings of the 7th ACM international conference on Multimedia, Orlando, Florida, USA, 303-311, 1999.

[19] M. G. CHRISTEL, A. G. HAUPTMANN, H. D. WACTLAR, et al. (2002). *Collages as dynamic summaries for news video*, Proceedings of the 10th ACM international conference on Multimedia, Juan Les-Pins, France, 561-569, Dec. 2002.

[20] B. CLARKSON, N. SAWHNEY and A. PENTLAND (1998). *Auditory Context Awareness in Wearable Computing*, Workshop on Perceptual User Interfaces, San Francisco, CA,, November 5-6, 1998.

[21] R. W. DEVAUL, S. J. SCHWARTZ and A. PENTLAND (2001) *MIThril : context-aware computing for daily life* http://www.media.mit.edu/wearables/mithril/MIThril.pdf.

[22] A. K. DEY and G. D. ABOWD (1999). *Towards a Better Understanding of Context and Context-Awareness*, Proceedings of the 3rd International Symposium on Wearable Computers, San Francisco, CA, pp. 21-28, October 20-21, 1999.

[23] A. K. DEY, M. FUTAKAWA, D. SALBER, et al. (1999). *The Conference Assistant: Combining Context-Awareness with Wearable Computing*, Proceedings of the 3rd International Symposium on Wearable Computers, San Francisco, CA, pp. 21-28, October 20-21, 1999.

[24] A. K. DEY (2001). *Understanding and Using Context*. Personal and Ubiquitous Computing Journal **5**(1): 4-7.

[25] R. JAIN (2003). *Experiential Computing*. Communications of the ACM **46**(7): 48-55.

[26] R. JAIN (2003). *Folk Computing*. Communications of the ACM **46**(4): 27-29.

[27] M. I. JORDAN (1999). Learning in graphical models. Cambridge, Mass., MIT Press.

[28] I. KANT (1920). Critique de la raison pure. Paris,, Flammarion.

[29] I. KANT, P. GUYER and A. W. WOOD (1998). Critique of pure reason. Cambridge ; New York, Cambridge University Press.

[30] M. KRUGER (1996). *Responsive Environments*. Theories and Documents of Contemporary Arts: A sourcebook for Artists' Writings, Berkeley CA, University of California Press, 1996.

[31] L. MANOVICH (2001). The language of new media. Cambridge, Mass., MIT Press.

[32] M. MCLUHAN (1994). Understanding media : the extensions of man. Cambridge, Mass., MIT Press.

[33] G. A. MILLER, R. BECKWITH, C. FELLBAUM, et al. (1993). *Introduction to WordNet: An On-line Lexical Database*. International Journal of Lexicography **3**(4): 235-244.

[34] M. R. NAPHADE and T. S. HUANG (2001). *Recognizing high-level audio-visual concepts using context*, Proc. IEEE Conference on Image Processing, 2001, Thessaloniki, Greece, pp. 46-49, Oct. 2001.

[35] M. R. NAPHADE and T. S. HUANG (2002). *Discovering recurrent events in video using unsupervised methods*, Proceedings IEEE International Conference on Image Processing. 2002., pp. 13-16, 2002.

[36] M. R. NAPHADE, C.-Y. LIN and J. R. SMITH (2002). *Learning semantic multimedia representations from a small set of examples*, Proc. IEEE International Conference on Multimedia and Expo, 2002, Lausanne, Switzerland, pp. 513-516, Aug. 2002.

[37] T. H. NELSON (1987). Computer lib/dream machines. Redmond, Wash., Tempus Books of Microsoft Press.

[38] R. PACKER and K. JORDAN (2001). Multimedia : from Wagner to virtual reality. New York, Norton.

[39] N. J. PAIK (1996). *Cybernated Art*. Theories and documents of Contemporary Art: A Sourcebook of Artists' Writings, Berkeley, CA, University of California Press, 1996.

[40] S. PFEIFFER, R. LIENHART, S. FISCHER, et al. (1996). *Abstracting Digital Movies Automatically.* Journal of Visual Communication and Image Representation **7**(4): 345-353.

[41] J. C. PLATT (2000). *AutoAlbum: clustering digital photographs using probabilistic model merging*, IEEE Workshop on Content-based Access of Image and Video Libraries, SC. USA, 96-100, Jun. 2000.

[42] J. RAMAMOORTHY, S. WONG and H. SUNDARAM (2003). *Dynamic Adaptive Visualizations*. Arts Media and Engineering Center, ASU, AME-TR-2003-03, 2003.

[43] N. SAWHNEY and C. SCHMANDT. (1999). *Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging*, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Pittsburgh, PA, May 15-20, 1999.

[44] B. N. T. SCHILIT, M.M. (1994). *Disseminating active map information to mobile hosts.* IEEE Network **8**(5): 22-32.

[45] B. SHEVADE and H. SUNDARAM (2003). *Vidya: An Experiential Annotation System*. Arts Media and Engineering Center, Arizona State University, AME-TR-2003-04, 2003.

[46] H. SRIDHARAN, H. SUNDARAM and T. RIKAKIS (2003). *Context, memory and Hyper-mediation in Experiential Systems*. Arts Media and Engineering Center, ASU, AME-TR-2003-02, 2003.

[47] H. SUNDARAM (2002). *Segmentation, structure detection and summarization of multimedia sequences*: xxviii, 331 leaves, bound., Thesis Ph D --Columbia University 2002.

[48] H. SUNDARAM and S.-F. CHANG (2002). *Video skims: taxonomies and an optimal generation framework*, Proc. of the International Conference on Image Processing 2002, Rochester, NY, 21-24, Sep. 2002.

[49] H. SUNDARAM, L. XIE and S.-F. CHANG (2002). *A utility framework for the automatic generation of audio-visual skims*, ACM Multimedia 2002, Juan-les-Pins, France, ACM Press, 189-198, Dec. 2002.

[50] H. SUNDARAM and T. RIKAKIS (2003). *An Introduction to Experiential Systems*. Arts Media and Engineering Center, ASU, AME-2003-01, 2003.

[51] S. UCHIHASHI, J. FOOTE, A. GIRGENSOHN, et al. (1999). *Video Manga: generating semantically meaningful video summaries*, Proceedings of the 7th ACM international conference on Multimedia, Orlando, Florida, USA, 383-392, 1999.

[52] B. VIOLA (1980). *Will there be condominiums in data space?* Video **80**(5).

[53] H. D. WACTLAR (1996). *The next generation electronic library—capturing the experience.* ACM Computing Surveys **28**(4es).

[54] H. D. WACTLAR, M. G. CHRISTEL, A. G. HAUPTMANN, et al. (1999). *Informedia Experience-on-Demand: capturing, integrating and communicating experiences across people, time and space.* ACM Computing Surveys **31**(2es).

[55] L. YE, C. HU, X. ZHU, et al. (2000). *A Unified Framework for Semantics and Feature Based Relevance Feedback in Image Retrieval Systems.*, In: Proc. ACM MM2000, 31-38,

[56] B. L. YEO and M. YEUNG (1998). *Classification, Simplification and Dynamic Visualization of Scene Transition Graphs for Video Browsing*, Proc. SPIE Storage and Retrieval of Image and Video Databases VI, San Jose CA, Feb. 1998., 1998.

[57] B.-L. YEO and B. LIU (1995). *Rapid scene analysis on compressed video.* IEEE Transactions on Circuits and Systems for Video Technology **5**(6): 533-544.

[58] M. YEUNG and B.-L. YEO (1996). *Time-constrained clustering for segmentation of video into story units*, Proceedings of the 13th International Conference on Pattern Recognition, 1996, 375-380, Aug. 1996.

[59] M. YEUNG and B.-L. YEO (1997). *Video visualization for compact presentation and fast browsing of pictorial content.* IEEE Transactions on Circuits and Systems for Video Technology **7**(5): 771-785.