

Storied Navigation

Toward Media Collection-Based Storytelling

Edward Yu-Te Shen

B.S EE, National Taiwan University, 2002

M.S. CSIE, National Taiwan University, 2004

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning,

in partial fulfillment of the requirements for the degree
of Master of Science in the Media Arts and Sciences

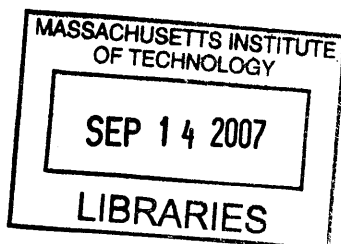
at the Massachusetts Institute of Technology
September 2007

© Massachusetts Institute of Technology, 2007
All Rights Reserved

Signature of Author
Program in Media Arts and Sciences
August 20, 2007

Certified by
Henry Lieberman
Research Scientist, Software Agents Group
MIT Media Arts and Sciences
Thesis Supervisor

Accepted by
Deb Roy
Chairperson
Department Committee on Graduate Students
Media Arts and Sciences



ROTCH

Storied Navigation

Toward Media Collection-Based Storytelling

Edward Yu-Te Shen

Thesis for Degree of
Master of Media Arts and Sciences at the
Massachusetts Institute of Technology

August 2007

Thesis Advisor
Henry Lieberman
Research Scientist
MIT Media Laboratory

Thesis Reader
Glorianna Davenport
Principle Research Associate
MIT Media Laboratory

Thesis Reader
Robb Moss
Rudolf Arnheim Lecturer on Filmmaking
Harvard University

Storied Navigation

Toward Media Collection-Based Storytelling

by Edward Yu-Te Shen

B.S EE, National Taiwan University, 2002

M.S. CSIE, National Taiwan University, 2004

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning on August 20, 2007

in partial fulfillment of the requirements for the degree
of Master of Science in the Media Arts and Sciences

Abstract

Life is filled with stories. Modern technologies enable us to document and share life events with various kinds of media, such as photos, videos, etc. But people still find it time-consuming to select and arrange media fragments to create coherent and engaging narratives.

This thesis proposes a novel storytelling system called Storied Navigation, which lets users assemble a sequence of video clips based on their roles in telling a story, rather than solely by explicit start and end times. Storied Navigation uses textual annotations expressed in unconstrained natural language, using parsing and Commonsense reasoning to deduce possible connections between the narrative intent of the storyteller, and descriptions of events and characters in the video.

It helps users increase their familiarity with a documentary video corpus. It helps them develop story threads by prompting them with recommendations of alternatives as well as possible continuations for each selected video clip. We view it as a promising first step towards transforming today's fragmented media production experience into an enjoyable, integrated storytelling activity.

Thesis Supervisor: Henry Lieberman

Title: Research Scientist of Media Arts and Sciences

Acknowledgement

Foremost, I would like to thank my three thesis readers: Henry Lieberman, Glorianna Davenport, and Robb Moss. This thesis would have been impossible to write without any of you. Henry, my advisor, you have been inspiring, encouraging and patient. From thinking about the big questions in Marvin Minsky's *Emotion Machine* to the hands-on experience of documentary filmmaking, you have supported me and guided me side by side on my way towards Artificial Intelligence and storytelling. You taught me how to appreciate the value of a piece of research work, offered me great freedom to build my thoughts however I wanted, and never hesitated to give me more time and wait with infinite patience. Thank you, Henry. You are the best advisor.

To me, Glorianna, my second advisor, talking to you is one of the greatest things in the Media Lab. Your words sometimes enlightened me with clues to answers, sometimes struck me with challenges, but they are all clearly imprinted on my mind. Your vision of corpus-based, computational documentary has provided the foundation of this thesis project; but your personal ways of telling stories and looking at the world have posed much more impact on me as a person. I am not sure whether we will only have four more months to work together after this thesis project is finished, but I truly wish our conversation can last ever after. Thank you Glorianna, for taking me as your own student.

Robb, my thesis reader at Harvard University, I cannot imagine what this thesis would be like right now, had I never had the opportunity to learn from you about every single step in filmmaking. It was honestly the best class I ever had in my life, because it gave me several things that are particularly meaningful and important to me, including this thesis. Your generosity to help and your insightful suggestions also made this thesis more solid and convincing. Thank you, Robb.

My mentors back in Taiwan: Ming Ouhyoung, Bing-Yu Chen, Yung-Yu Chuang, Rung-Huei Liang, Jane Yung-Jen Hsu, Ja-Ling Wu, and Hao-hua Chu. You gave me great support and encouraged me to pursue a bigger challenge. I would not have been able to come to the Media Lab and finish this thesis without your support. Thank you all.

I would also like to thank another mentor of mine, Hugh Herr, who posed great impact on this thesis. If, to any extent, I have become stronger or braver in either research or life, Hugh, it is because of you. And, Barbara Barry, my mentor and friend, thank you for giving me so many great suggestions and warm encouragement.

Other members of the media lab also gave me great support. Linda Peterson and Tesha Myers,

thank you for helping me in the process of making *Life. Resarch*. Mary Heckbert, Betty Lou McClanahan, and Gigi Shafer, thank you for supporting me with all sorts of things that I needed. Paula Aguilera, Jacqueline Karaaslanian, and Brian Wallenmeyer, it has been a great experience to work on storytelling with you. My friends in the Media Lab: Seth, Paulina, Hyun, Sam, Pranav, Dustin, Jaewoo, Paul, Kevin, Ian, Moin, Kelly, David, Rob, Catherine, and many others, it has been an invaluable experience to do brainstorming with you. It was fun, and I learned a lot.

My friends who have been there with me during the two-year journey: James, Francis, Daniel, Elissa, Chao-Chi, Wu-Hsi, Jackie, Muffin, Yu-Chiao, Cindy, Helen, Brook, Anna, Carrie, Elisa, Taco, Emily, George, Michael, Shunsuke, Shirley, Ying-Zi, Willy, Kuba, Cheng-Yang, and Jing-Li, and those who have been supporting me over the Pacific Ocean: Zephyr, Brian, Ray, Pan, Xenia, Drake, Sula, Jenhuit, Charity, and many more. Your friendship is my real treasure.

I would like to thank Epoch Foundation in Taiwan, lead by Josephine Chao, for always being there and supportive through all these years. Also the National Science Council of Taiwan, the Department of Computer Science and Information Engineering and Military Education Office of National Taiwan University, and the Military Service Department of Taipei City Government, thank you for all sorts of support too.

A million thanks to Henry Lieberman, my advisor, for proofreading my thesis word by word. I really appreciate it. Also many thanks to Cindy, Grace, and Yu-Chiao, for helping me out at this very last moment; my uncle Cheng-Zhong, thank you for praying for my thesis through all the nights; and my mentor and friend, John Thomas, thank you for all the things you have taught me, and for encouraging me even at the time of facing your own biggest challenge.

Mom, Dad and my two sisters, Grace and Alice – the ones that I love the most in the world. If I made any achievement, it was because of the strength that you have given me. Thank you all. And, thanks Mom, I didn't speak out on the phone, but I love you with no doubt.

And, with all my heart, Elissa Huang, thank you for being generous, encouraging, and, most important of all, understanding. You are the one who made me free.

Finally, I wish to dedicate this thesis to James Teng, who started to help and to guide me from my very first moment of applying for the Media Lab, and never stopped. Things are hard to put in short words, but, I've been always appreciative.

Table of Content

I Introduction.....	19
1.1 Motivation.....	19
1.2 Problem Definition.....	20
1.3 Example Scenario.....	22
1.4 Contribution & Generalization.....	23
1.5 Roadmap.....	24
II Theory & Rationale.....	25
2.1 Indexing Stories.....	25
2.2 Storied “Ways to Think” in <i>Life. Research</i>	27
2.3 Commonsense Computing.....	35
2.3.1 OMCS: A Knowledge Base for Common Sense.....	36
2.3.2 ConceptNet: A Common Sense Semantic Network.....	36
2.3.3 WordNet & MontyLingua for Natural Language Processing.....	37
2.4 Story Analogy & Story Feature Computation.....	38
2.4.1 Extracting Story Features.....	38
2.4.2 Extracting Schank’s Story Indices.....	39
2.4.3 Using the Features: Story Analogy.....	40
2.5 Summary.....	44
III System Design & Implementation.....	45
3.1 Interaction Design.....	45
3.2 Algorithm.....	48
3.2.1 Parsing the Story Descriptions.....	48
3.2.2 Annotating Video Sequences.....	54
3.2.3 Making Stories by Typing Story Descriptions.....	55
3.2.4 Searching Sequences by Story Features.....	58
3.2.5 Recommendations: “Find Similar Alternative” & “What’s Next?”.....	58
3.3 Implementation.....	59
3.4 Summary.....	59
IV Using the Storied Navigation System	61
4.1 Video Annotation.....	61

4.2 Storytelling.....	64
4.3 Summary.....	70
V Observation & Evaluation.....	71
5.1 Study 1: Using the System to Make Video Stories.....	71
5.1.1 Study 1: Design & Setting.....	71
5.1.2 Study 1: Observation & Results.....	74
5.1.2.1 Study 1: Subject 1.....	74
5.1.2.2 Study 1: Subject 2.....	77
5.1.3 Study 1: Summary.....	82
5.2 Study 2: Using the System to Browse Videos.....	82
5.2.1 Study 2: Design & Setting.....	82
5.2.2 Study 2: Observation & Results.....	84
5.2.3 Study 2: Summary.....	89
5.3 Statistics of the Questionnaires.....	90
5.4 Summary.....	91
VI Discussions.....	93
6.1 Syuzhet, Fabula, and the Story Path.....	93
6.2 The “Accompanied Editor”.....	95
6.3 Goal.....	95
6.4 Lessons, Inspirations, Reflections.....	96
6.5 The “Democratic” Interactive storytelling.....	98
VII Related Work.....	101
7.1 Commonsense Computing.....	101
7.2 Interactive Storytelling.....	101
7.2.1 Media Collection-Based Storytelling.....	101
7.2.2 Character-Based Storytelling.....	103
7.3 Applications of or Discussions about Story Representation & Narrative Theory.....	104
7.4 Interface Design for Browsing/Annotating Media Collection.....	106
7.5 Video Retrieval based on Automated Annotation.....	107
7.6 Other Work about Video Semantics.....	108
7.7 Automated Video Editing.....	109
7.7.1 Automated Documentary Video Editing.....	109
7.7.2 Video Summarization.....	109

VIII Conclusion.....	111
IX Conclusion.....	113

List of Tables

Table 1-1.....	20
Table 2-1.....	29
Table 2-2.....	32
Table 2-3.....	40
Table 2-4.....	41
Table 2-5.....	42
Table 2-6.....	43
Table 2-6.....	44
Table 3-1.....	45
Table 5-1.....	78
Table 5-2.....	86
Table 5-3.....	86
Table 5-4.....	87
Table 5-5.....	91

List of Figures

Figure 1-1.....	22
Figure 1-2.....	23
Figure 2-1.....	37
Figure 3-1.....	46
Figure 3-2.....	46
Figure 3-3.....	49
Figure 3-4.....	50
Figure 3-5.....	51
Figure 3-6 (a).....	52
Figure 3-6 (b).....	53
Figure 3-7.....	55
Figure 3-8.....	56
Figure 4-1.....	61
Figure 4-2.....	62
Figure 4-3.....	63
Figure 4-4.....	63
Figure 4-5.....	64
Figure 4-6.....	65
Figure 4-7.....	66
Figure 4-8.....	66
Figure 4-9.....	67
Figure 4-10.....	67
Figure 4-11.....	68
Figure 4-12.....	69
Figure 4-13.....	69
Figure 4-14.....	70
Figure 5-1.....	73
Figure 5-2.....	90

I Introduction

"Life is pregnant with stories " – R. Kearney, 2002.

1.1 Motivation

Today, recording and sharing real life events with video has become more and more popular on repository websites like YouTube™ [21]. From TV programs like news, sports, entertainment, to personal documentary such as travels, birthday parties, graduations ceremonies, etc., videos are recorded and uploaded because they carry particular meanings, especially to viewers/owners who are involved in the documented events or contexts. Regarding these videos as small "bits" of potential bigger stories (e.g., the review of a child's growth process or of a nation's progression to democracy), the variety of and complex relationships between these videos suggest that countless stories could be formulated by taking multiple progressive paths through the collection. Under this scenario, the story that a viewer gets would depend on his/her selection according to his/her reflection upon each video, his/her perspective, background, purposes, etc. As different viewers view or juxtapose the videos in different sequential orders, different meanings emerge.

Nevertheless, finding and sequencing these video clips is a labor-intensive task, since there is no mechanism that facilitates (re)discovery of events or facts in those videos in a *storied* way. That is, the viewers cannot navigate the video corpus in a progression that highlights interesting/memorable relationships (e.g., chronology, causality, etc) between the individual clips. Browsing tools such as keyword search are not rich enough for a storied navigating experience because, in my perspective, they are not designed for performing any of the humans' "ways of thinking" [15] for continuing a story, introduced in Chapter 2. To insure coherent stories and more fluent, enjoyable viewing experience, we need to go beyond keyword search.

In this thesis, we propose *Storied Navigation*, a methodology for improving *story composition*, which I define as the meaningful temporal ordering of video sequences. As users navigate throughout the *story world* – a video repository – they will encounter *story segments* – video sequences – one after another, in a meaningful storied way; and *story threads* – series of video sequences, will be formed by the navigation path. Since they can freely decide how the stories evolve, the users will be no longer limited by static, linear storylines as in conventional video viewing activities.

The target users of Storied Navigation are the creative, entrepreneurial amateurs who have

massive amount of energy and ideas but not much experience, and who want to proactively make new stories, particularly sharable content on the Internet. These users are the so-called “prosumers”, or professional consumers. What they need is a tool that will give them a set of recommendations to choose from as they navigate through the large database of video sequences, based on their own likes, tastes, or story needs. Table 1-1 shows the comparison between professional editors, amateur or home video makers, and the target users of storied navigation.

Table 1-1: Comparison between Different Users of Video Storytelling

	Storytelling Experience	Skill Level	Story Making Creativity	Stories They Wish to Tell	Example Tools that They Need
Professional Editors	A lot	High	Great	Complex	Final Cut Pro, Adobe Premiere, or other professional software
Amateur, Home Video Makers, or Viewers	A little	Low	Vary	Simple	iMovie, Windows Movie Maker
Professional Consumers (Prosumers)	Little or Medium	Low or Medium	Great	Complex	Storied Navigation

Both the conventional “viewers” and “filmmakers” can benefit from in this novel scenario too. Viewers who currently sit in front of the screen and passively consume the stories, can become participative and enjoy various possible story paths by following the guidance of the system’s recommendations; Professional filmmakers who do cut-and-paste with the images using modern editing software (e.g. Adobe Premiere, Final Cut Pro, etc.) can use Storied Navigation as a tool that assists the understanding of their materials, as well as the development of possible story threads. This thesis project is not immediately aimed at fully supporting Storied Navigation of online repositories like YouTube, which presently consists of large numbers of unconnected, unannotated, small clips. However, as YouTube publishers and viewers gather increasingly sophisticated personal corpora of material and attempt increasingly sophisticated stories, they will move closer to the underlying authoring mechanisms that we study here.

1.2 Problem Definition

In making a movie, an editor *arranges sequences into a temporal viewing order* based on his/her *profound understanding of the meaning of each video sequence*, such that the audience will be presented with a well-designed story. In the process of viewing videos online, there is no editor to make the stories smooth, interesting, and understandable. Thus, we wish to design a system which facilitates the users similarly to the experienced editor that exhibits these two capabilities.

There are two major problems in this thesis, 1) “How can a system help users to gain familiarity with the materials stored in the corpus?”, and 2) “How can a system help users to develop their story threads?” Below, we discuss these two problems more deeply.

Video is opaque. One cannot understand what a video conveys unless he/she watches the video or someone else describes it to him/her. Since today’s computers cannot yet understand the meaning of a video by “watching” it (processing its visual/aural signals), they have to rely on a piece of summarized description provided by a human. Often, this text describes events, characters, and locations of a scene, which do not necessarily represent how that scene might play a role in a larger or complete story. Thus, in order to help users get familiar with the corpus, the system needs to match descriptions of the role that a scene might play in a bigger story with descriptions of what is happening in the scene. If we assume that both the users’ input and the videos’ descriptions are free text, we can re-formulate the first problem as:

How can a system match textual descriptions of the narrative of a story, with textual descriptions of particular video scenes, in order to help the user understand the contents of a video corpus?

Second, in making a film, an editor generates the story flow based on what he/she knows about the context of the real world and story models. If we assume that the user can initiate the direction of the next story segment based on his/her curiosity, what is left for the system to do is to leverage knowledge about the video corpus and common sense to provide video sequences that substantiate and foster the initiated direction. Thus, the system will not be solving a traditional search problem, since the users may not be able to anticipate the sequences that they are looking for, and it is not trivial to clearly determine whether or not a video is a correct result for the input. Since the system may also record the relationships between videos or between videos and users’ narrative during the interaction, the second problem can be re-formulated as:

How can a system find story segments whose textual descriptions are related to the input and that are potentially interesting to the users, based on 1) common sense knowledge that is available to the system, 2) the videos’ textual descriptions, and 3) the user-contributed information about the interrelations between the videos?

In Chapter 3, we introduce our idea of solving the problems by proposing a hybrid scenario that integrates browsing and editing a video corpus. It allows users to navigate by asking for an arbitrary clip, inputting *story descriptions* – which I will define in Chapter 3 – in English, or selecting system-prompted clips from “Find Similar Alternatives” or “What’s Next?” recommendations. Thus, the aforementioned two problems can be reformulated into the following sub-problems: a) From a

piece of story description, how do we extract features that are potentially useful in humans' "ways of thinking" in story continuation activities? b) Using the extracted features, how do we design a metric that measures the distance between two videos' semantic meanings, such that the system can recommend similar alternatives for a specified clip?, and c) Similarly to b), for a set of candidate videos, how do we design a metric that measures their appropriateness of one following another in temporal order, such that a coherent, inspiring, and intriguing story flow can be produced? The main ideas of the proposed solutions to these problems originate from the analysis of my documentary film, *Life. Research.* in Chapter 2, which serves as the theoretical foundation of this thesis document.

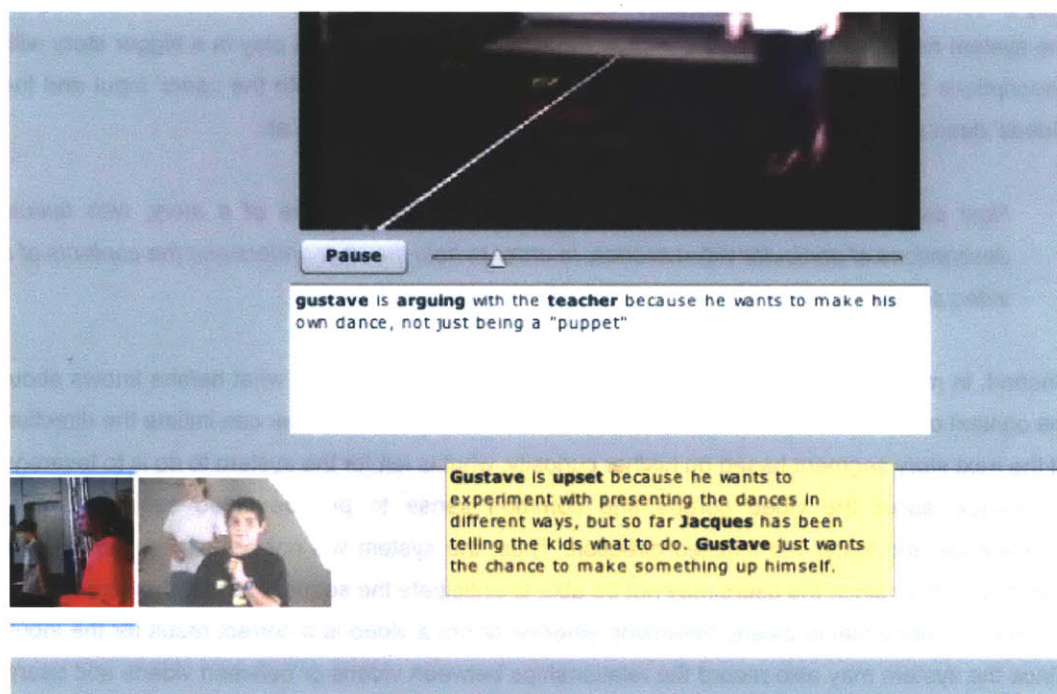


Figure 1-1: Example Scenario of Using the Storied Navigation System (a)

1.3 Example Scenario

To give a clearer picture of what this kind of solution might be, below is an example scenario of using the Storied Navigation system. Figure 1-1 shows the interface of the Storied Navigation, where a user types a sentence and the system responds with a few video sequences in the timeline. The sentence that the user types is, "gustave is arguing with his teacher, because he wants to make his own dance, not just being a 'puppet'", where the annotation of one of the retrieved videos is "Gustave is upset because he wants to experiment with presenting the dances in different ways, but so far Jacques has been telling the kids what to do....", as shown in the

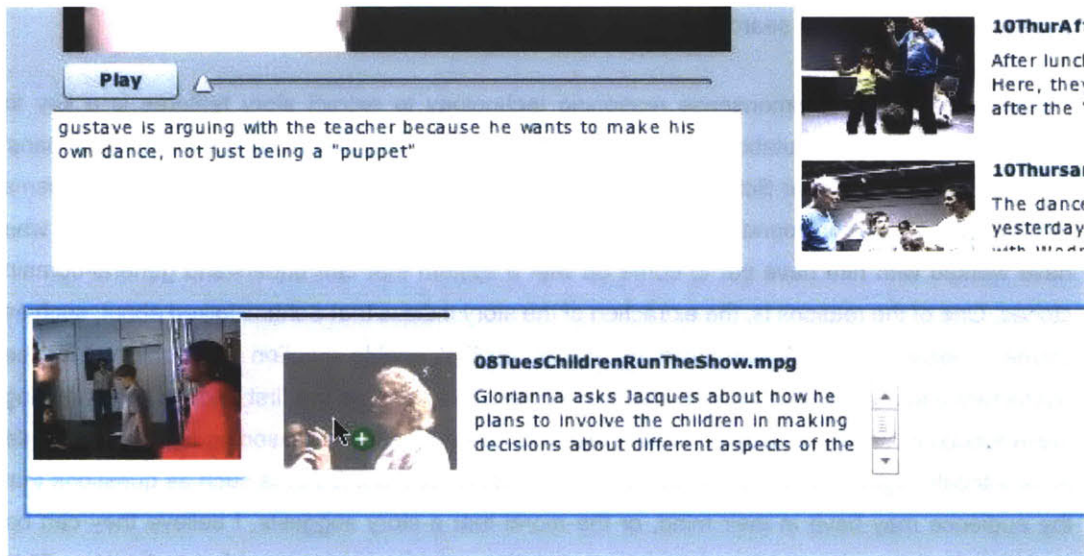


Figure 1-2: Example Scenario of Using the Storied Navigation System (b)

yellow box. From the interface we can see that, what the system does is that it tries to understand the matches between the videos' annotations with the user's input, and responds with a series of sequences that form a story all together.

Then, if the user uses "What's Next?" to ask for recommendation about the following videos based on the "theme" criterion, one of the best results returned by the system will be a sequence where Jacques was answering questions about whether he would allow the children to make their own decisions, as shown in Figure 1-2. The juxtaposition between these two sequences is somewhat interesting, because to certain degree it shows a contradiction between what Jacques said and what actually he did. This example suggests how a *storied* navigation experience may be. The users will not be simply performing series of keyword searches that are independent from one another. Instead, the users will be navigating in a continuous flow of their interested characters, emotions, events, topics, etc, by the system's support. In Chapter 4 I will give a more complete presentation of the functions provided in the Storied Navigation system.

1.4 Contribution & Generalization

This thesis proposes the first methodology that 1) extracts as many as five types of story features from free text stories (character/role, emotion, location, theme, date) by using commonsense reasoning technology; and 2) exhibits a certain degree of human-like intelligence making suggestions for story construction. Various kinds of other useful applications are possible using similar methodology, such as textual story editors, collection-based text-to-animation systems, a storytelling companion for fostering children's storytelling capabilities, search systems for FAQ

posts personalized to the searchers' needs, and many more.

I believe that using commonsense reasoning technology to extract story features is a key to understanding how computational machinery might be able to emulate the many kinds of humans' *storied* intelligence. Roger Schank has written several books on story features and mechanisms that might be used to ground an Artificial Intelligent system; however, to date researchers who have worked with him have yet to come up with a system that can understand general-domain stories. One of the reasons is, the extraction of the story indices that Schank talked about, such as theme or lesson, requires computational resources that enable emotion understanding of the characters and analogy making between two stories. This thesis is the first attempt of performing these functionalities by employing such a resource – commonsense reasoning technology. While more intricate algorithms will be required to extract more abstract features such as questions that the audience may have in their mind, or the moral that a story suggests, I believe they can be achieved based on the same methodology of using commonsense reasoning technology. This methodology is, I think, the most valuable contribution of this thesis.

1.5 Roadmap

The thesis is organized in 9 chapters. Chapter 2 introduces the main theory and rationale of this thesis, including a list of “ways to think” in making video stories derived from the analysis of my documentary film, *Life. Research*, and an introduction of the related research of Roger Schank, which has a similar philosophy with my film analysis. Based on these thoughts, I introduce how the system is designed and implemented in Chapter 3. Chapter 4 presents a more thorough presentation of how the system can be used, and Chapter 5 reports the observation and evaluation of the two conducted user studies. In Chapter 6 I discuss about various thoughts that are related to filmmaking and other topics, and Chapter gives a introduction of the related literature. Finally, I conclude the thesis in Chapter 8, followed by the references in Chapter 9.

II Theory & Rationale

"...the indices for retrieving a story would be exactly those that were used to represent it. In this way, we could be reminded of a new story by an old story simply by storing the new story in terms of the same elements we have used to understand both stories, namely the themes, goals, plans, and such that characterized what was going on in each story." – R. Schank, 1991.

This chapter introduces the theoretical foundation of the Storied Navigation system. First I will go through Roger Schank's theory on indexing stories, which provides part of the theoretical foundation of this thesis. To understand more about storytelling using the media of documentary video, I will also discuss about a film "Life. Research.", which I made prior to the system design. There are 16 scenes in this short film, each of which performs a particular function in the whole story. In section 2.2, I will present an analysis of these 16 scenes, followed by a list of "ways to think" for choosing consecutive video sequences in making this film. Finally, by introducing the commonsense computing tools, I will discuss about how commonsense computing may be useful in achieving these ways of thinking.

2.1 Indexing Stories

In this paper, a *story* is defined as a sequence of facts about a set of story elements including characters, objects, locations, events, time, and the interrelationships between different elements. Consider the sentence, "So what we're gonna do is to put this stuff on the web, such that our customers will be able to check their own data", "what we're gonna do is to put this stuff on the web", and "such that our customers will be able to check their own data" are the two facts; "we", "what we're going to do", "this stuff", "the web", "our customers", "their own data" are the story elements; and "is", "put", "on", "such that", "will be able to check" are the relationships. A question is not a story, but it certainly is a question about whether a story is true. An example like, "Can we go out and have some spaghetti tonight?" is a question about whether the story "We can go out and have some spaghetti tonight" holds true.

On the other hand, a *story description* is defined as a story written as one or more English sentences, whose characters are all third-person nouns, proper names, or pronouns, unless the character appears in a dialog. Hence, the sentence "You should eat breakfast." is not a story description, whereas "Chris answered Paul, 'I don't want to go to the dinner'" is. Such a definition

will simplify the task for a computer system, because it will not need to spend much time specifying which person “I” refers to, or which person “you” refers to.

I ground my theory on the research work of Roger Schank, because his thoughts matches my personal experience and observation in filmmaking very well, as opposed to other theories such as *narratology*, which talks about narrative using terminology including *syuzhet*, *fabula*, etc [65]. According to Roger Schank [17], “A mind must be able to find what it needs to find, and it must know that it has found it.” In order to impose the computer with the capability of processing these stories, the first step is to teach it how to find the stories, as how humans do. Schank argues that, there are too many stories that are stored in our brains, and if we don’t have a good way to locate the stories that we wish to find, there is no way we can utilize them. Thus, a good *indexing* strategy is important to store, retrieve the stories.

Finding this good strategy for a computer system is not an easy task, however, if it aims at finding interesting stories for human users. This reason is, human beings index stories much more profoundly than simply using the contained elements in the story text, such as characters, time, and location. Below is an example given in his book [17]:

- i. “Trying to prove yourself may cause you to do things you don't want to so you can appear cool”
- ii. “Sometimes you act irrationally in a group when you wouldn't if you were alone”

A human reader can understand that these two little stories are somehow similar to each other, even though they don’t necessarily share the same characters, objects, places, events, time, or interrelationships. This is because, according to Schank, human beings use more high-level story indices in reasoning and manipulating stories. He proposed using topics or themes, goals, results, lessons or observations as the indices. By listing these indices for the sentences above, we can derive an interesting comparison between the two examples:

- i. “Trying to prove yourself may cause you to do things you don't want to so you can appear cool”
Topic: group acceptance | Goal: appear cool | Actual Results: feel lousy | Lesson: Be yourself
- ii. “Sometimes you act irrationally in a group when you wouldn't if you were alone”
Topic: group acceptance | Goal: appear cool | Actual Results: feel lousy | Lesson: Be yourself

As the readers can see, although it is very difficult to tell from the textual appearance that the two sentences are analogous in many ways, surprisingly the indices are almost identical. In other

words, based on Schank's theory, two analogous stories tend to share similar topics or themes, goals, results, and lessons or observations. If a system could extract these indices from story text, it would be able to search software documents in a more intelligent way.

A good way of indexing stories is useful not only for fast and accurate storage and retrieval, however, but for effective and efficient storied thinking as well. Schank listed a set of questions that a good theory of mind must be able to explain [17]:

- a) How are stories indexed in our mind?
- b) What are the representations for the stories stored in our mind?
- c) How are memories retrieved using these indices?
- d) How can different stories be matched based on their indices?
- e) According to what does our mind decide to match new stories with old ones?

That is, story indices are used by humans to represent what a story means to us, and to compare, relate, and analogize one story to another. So if a computational system can extract and utilize these indices, it is more likely for the system to manipulate the stories in a more interesting, intelligent, *human* way, than conventional video browsing facilities such as keyword search. In section 2.3, we will discuss how we can possibly develop a practical method to build a machine that can extract and process these indices from plain story descriptions.

In summary, a Storied Navigation needs a way to index the textual stories that are used to describe the stored video sequences. And the way it indexes these stories shall be as close to Schank's indices (themes, goals, results, lessons) as possible, in order to behave intelligently by finding, matching, performing analogy between stories, and providing interesting recommendations to the human users.

2.2 Storied "Ways to Think" in *Life. Research.*






Although the two given short examples in the previous section fall into the definition of stories, they are actually "bigger" stories, i.e. they are more abstract, condensed than most of the stories that would appear as descriptions for 1-2 minute video clips in a Storied Navigation system. To think about the indexing problem with a more concrete example which is closer to the target media – video – in the system to build, in this section I try to induce a set of indices by analyzing a film.







Life. Research. is a 15-minute documentary film that I made in early 2007, cut from a collection of 12-hour raw footages. It is a story about Hugh Herr, who is a double amputee and now a professor at the MIT Media Laboratory. Step by step, the story reveals Herr's research on artificial limbs, his own physical inconvenience, his interaction with sponsors, students, family members, and finally





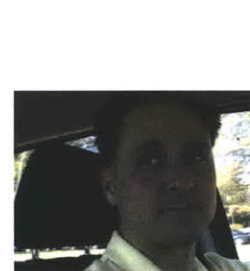
his ambition, the difficulties he has been through, and, implicitly, his philosophy of life. In Table 2-1, the analyses for the scenes are shown. Each row depicts the information of each scene, including a thumbnail, the scene's name, a piece of description of what happened in the scene, some key dialogs appeared in the scene, and, finally, the function that this video sequence serve in the whole story.

From Table 2-1, we find that every single scene performs its own function in the whole story. Some raise questions, others answer previously-raised questions, and still others give surprises, conclusion, or any kind of new information. Like constructing a building, every new scene or building block performs its function based on the preceding ones. That is, a video clip will not function the same way as in Table 2-1 if it is placed in a different sequential order.

Table 2-1: Scene Analysis in “Life. Research.”







Name	Thumbnail	Scene Description	Key Dialog	Scene Function
Introductory scene		Establishing shot: Stills of Hugh's lab and inventions.		<ol style="list-style-type: none"> 1) Beginning of the story. 2) Getting the audience's attention
Title scene		The film's title		<ol style="list-style-type: none"> 1) Showing the title of the film. 2) Raising questions: "What is it all about?" "What is the relationship between life and research?"
Lab scene 1		Hugh giving a presentation of the research in the lab	<p>"Blurring the boundary between humans and devices"</p> <p>"Two key applications: rehabilitation and augmentation"</p>	<ol style="list-style-type: none"> 1) Introducing the main character 2) Impressing the audience with the technology
Lab scene 2		Hugh demonstrating the technology on his own legs and answering a few questions.	"I lost my legs in 1982"	<ol style="list-style-type: none"> 1) Showing that Hugh himself is an amputee, and he was not born with this condition 2) Raising questions: "What was the reason for his amputation?" "Is that related to why he is working in this field? "
Ambition		Hugh talking about his plan for commercialization as well as the future goal while answering a visitor's questions.	"Maybe do a \$100-like, so that everybody in the world can benefit from high technology"	Showing Hugh's ambition and his willingness of helping people all over the world.








Courage in career decisions 1		Hugh's student Sam relates how Hugh convinced him to change his field	"Build something useful to a disabled person"	Continuingly, showing Hugh's aggressiveness of asking people to join realizing his goal, and restate his will of helping people with disabilities.
Courage in career decisions 2		Hugh and another student Edward conversing in the car about Edward's career	"Have someone you trust and admire to make the decision for you... Isn't that weak?" "You may actually want to get a PhD but you're too frightened to take actions".	Showing Hugh's recognition of the difficulty of making the right choice, being responsible for it, and taking actions
Background 1(a)		Introductory of Hugh's wife and children.		Showing Hugh's family life and how that reveals his devotion to the ones that he loves
Background 1(b)		Hugh's wife explaining how kids enjoy helping him put on his legs.	"It's like...you forgot these"	1) Showing how close Hugh is with his family members, and the happiness they share with each other 2) The audience also starts to see how his research and personal life are connected on a daily basis
Surprising Fact		Hugh running around a lake on a pair of artificial legs.		Surprising the audience and raising questions like, "How can he run so fast?"
Explicit statement of theme		Edward being amazed that Hugh can run through the forest. Edward and Hugh talking about research.	"So, you actually get your research from every single moment." "Research is life. Life is research" "I wanted to be an athlete "	1) The highest emotional point in the whole story 2) Explaining how Hugh can run like that 3) Implicitly posing a hint: "There might be some reasons for him to change his direction"




<p>Leg Changing scene</p>		<p>Hugh changing his artificial legs after he ran</p>		<ol style="list-style-type: none"> 1) While being a peaceful scene, it inspires the audience with the idea of how Hugh strived throughout the years, both mentally and physically 2) And it also raises a fundamental question in the audience's minds: "What is the cause that made him who he is today?"
<p>Background 2</p>		<p>Textual narration of accident details</p>	<p>"Hugh's legs were frostbitten" "One died in the rescue attempt".</p>	<p>Giving a clear explanation of what happened in the past, and also answer to many of the previous questions</p>
<p>Controversy</p>		<p>Magazine editor being critical of Hugh's athletic accomplishment as "artificial". Reader jumps to Hugh's defense.</p>	<p>"While impressive, it doesn't count as a legitimate free ascent" "Your whining is nauseating".</p>	<p>The only scene where Hugh shows great relief: Finally someone fights back for him.</p>
<p>Background 3</p>		<p>Patricia tells the history of Hugh as a fearless climber and self-taught researcher</p>	<p>"Everything after high school was completely self-taught" "He decided that he wanted to learn, and he learned"</p>	<p>Explaining that Hugh's motivation came completely from his experience rather than from academic interest. Research is driven by life.</p>
<p>Courage in career decisions 3</p>		<p>Reprise of car discussion between Hugh and Edward about PhD.</p>	<p>"You gotta do what's right"</p>	<ol style="list-style-type: none"> 1) Showing Hugh's belief of "what's right" – contributing to other people by overcoming the fear in one's heart 2) Resonating and making an analogy between Edward's dilemma and Hugh's own decision making experience in the past.

If we focus on why a clip is placed at its current position and how the transition between two consecutive clips "works," i.e. makes sense to the audience, we can derive another table below:

Table 2-2: Transition Analysis in “Life. Research.”

Name	Thumbnail	Scene Description	The Reason Why the Scene is “Here”
Introductory scene		Establishing shot: Stills of Hugh's lab and inventions.	Establishing shots are suitable for beginning
Title scene		The film's title	The title has to appear early in the film
Lab scene 1		Hugh giving a presentation of the research in the lab	It sets the context (a research lab, prosthesis research) and introduces the main character (Hugh), which should also be early in the film, since the rest of the clips all have to be introduced latter than this one.
Lab scene 2		Hugh demonstrating the technology on his own legs and answering a few questions.	It shares a similar visual image (somebody wearing artificial limbs walking on the stairs), and similar context (research lab, prosthesis research) with the previous clip. It also shows that Hugh himself is an amputee, and, more importantly, because of some accident, which has to be introduced at this time for the rest of the material to follow
Ambition		Hugh talking about his plan for commercialization as well as the future goal while answering a visitor's questions.	Again, it shows similar context with the previous clip. It also shows the <i>ambition</i> and <i>goal</i> that Hugh has, which will lead to and resonate with the <i>reason</i> why he is doing this in the forthcoming clips.
Courage in career decisions 1		Hugh's student Sam relates how Hugh convinced him to change his field	It shares and strengthens the idea that Hugh wants to make the prosthesis for other people's good. That Hugh encourages other people to change directions because of “what's right” also starts to resonate with the rest of the film from here.

Courage in career decisions 2		Hugh and another student Edward conversing in the car about Edward's career	It is analogous to the previous scene in that both scenes give a story about Hugh advising a student in a career decision making situation. As a real-life event, it also makes the previous interview livelier in front of the audience.
Background 1(a)		Introductory of Hugh's wife and children.	It sets a new context: "family". The film has to turn to the family side of Hugh after all the preceding scenes, in order to give the audience a more intimate, more humane feel of this person.
Background 1(b)		Hugh's wife explaining how kids enjoy helping him put on his legs.	It follows the family context set in the previous clip. It also strengthens the concept that Hugh and his family members love each other very much.
Surprising Fact		Hugh running around a lake on a pair of artificial legs.	While the rhythm of the story gradually gets calm and slow, it surprises the audience with a stunning shot. It also prepares the appearance of the next coming clip by setting the context and lifting the emotion of the story
Explicit statement of theme		Edward being amazed that Hugh can run through the forest. Edward and Hugh talking about research.	It points out how the film got its title, which should be about where it is (about 2/3 of the overall time length), and reaches the highest emotional point. The dialog "I wanted to be an athlete" also prepares the appearance of the following magazine scene.
Leg Changing scene		Hugh changing his artificial legs after he ran	It shares the same context with the previous two clips. It also gives the audience more time and space to digest what they just saw, and to think about what he really meant by "Research is life, and life is research."
Background 2		Textual narration of accident details	It answers all the questions that the audience may ask up to now. It actually also has to appear around this time (about 3/4 of the overall time length), such that the audience will have time to reconsider how he is right now, based on the understanding of the tragedy.

Controversy		Magazine editor being critical of Hugh's athletic accomplishment as "artificial". Reader jumps to Hugh's defense.	It has to be after the introduction of the accident, because the audience has to be aware of the accident and his rock climbing ability before watching this scene. And it serves as a transition from the sad emotion to a backed-up, relieved mental state.
Background 3		Patricia tells the history of Hugh as a fearless climber and self-taught researcher	It brings the audience back to Hugh's childhood, and allows them to stand at a different point of view to look at who he is today. It functions much more strongly, since the audience knows all the previously-introduced information.
Courage in career decisions 3		Reprise of car discussion between Hugh and Edward about PhD.	It comes later than "Courage in career decision 2", so that the audience understands the context of the conversation. It comes right after "Background 3" because the idea of "Made the decision for more people's benefit, and do it once the decision is made" can resonate strongly. Its final sentence "You gotta do what's right" also serves as a great ending for the whole story.

Thus, we can conclude that the way I used to arrange these video clips into a film depends on eleven different criteria. That is, a sequence can be chosen to follow the previous sequence if:

- a) it reveals some information that has to be delivered to the audience at this exact point, or
- b) it exhibits an opening, ending, or other kinds of function to the whole structure, or
- c) it establishes a new context for a new story part to follow, or
- d) it strengthens, extends the story, or elevates the current story to a different level, or
- e) it brings conflicts, surprises, complex, puzzles, or other kinds of unexpected, emotionally-strong facts and to the existing story, or
- f) it provides explanations to the questions in the audiences' minds such that the story can proceed, or
- g) it presents related meanings to the existing story from a different point of view, context, with a different strategy, or in a different form, or
- h) it shares analogous concepts (relationships, attitude, interaction, etc) with or relates conceptually to the previous sequence, or
- i) it is continued from the previous sequence in terms of time, location, character, event

Referencing Minsky's "Emotion Machine" [15], I call these criteria humans' different "ways to think" in the context of story continuation. It may or may not be a complete list for all sorts of humans' storied thinking ways, yet the completeness of this list is not what this thesis's focus. To make a more condense list of these "ways to think", the features that can facilitate a system to recommend

adjacent sequences would be:

Content of the video sequence:

- 1) The characters (d, e, f, g, h, i)
- 2) Interactions of the characters (c, d, f, g, h)
- 3) Emotion of the characters (d, e, f, g, h)
- 4) Time (c, l, f)
- 5) Location (c, l, f)

Function of the video sequence:

- 6) Whether playing an opening, establishing, ending, or other kinds of functional role in the whole structure (c, b)
- 7) Whether it has to appear at a certain point in the story flow (a)

In other words, to determine whether criteria a) (“this video sequence reveals some information that has to be delivered to the audience at this exact point”) holds true, the system requires the information of 6) whether this sequence has to appear at a certain point in the story flow, which may need to be specified by the human users. As another example, to determine whether criteria d) (“this video sequence strengthens, extends the story, or elevates the current story to a different level”) holds true, the system requires the information of 1) the characters, 2) the characters’ interactions, and 3) the characters’ emotions. Once the system has these three types of information, it may be able to analyze whether the selected video sequence shares the same or analogous pairs of characters with the preceding one, whether its emotions are similar, or even stronger, to those in the preceding one, etc. The attempt of performing analogy will be introduced in section 2.3.4.

To distinguish these criteria from the above ways to think and the story indices proposed by Roger Schank, I will refer to these criteria as “story features” throughout this thesis. In the later section, we will show how we extract the story features from free-text stories by considering the capability of today’s computer technology.

2.3 Commonsense Computing

The tool that I use to extract and utilize the story features in the Storied Navigation system is called the commonsense computing technology. This section explains why these commonsense computing tools are suitable for this task, by introducing the projects, together with their respective strengths. Then, in the next section, I will introduce how I determined the story features actually used in the system, based on the tools’ capabilities.

2.3.1 OMCS: A Knowledge Base for Common Sense

The first reason why commonsense computing technology has to be leveraged, is that our system needs to have a wealth of everyday knowledge in order to understand all story descriptions to certain degree, such that it can respond the human users in a useful way [35]. In other words, a computer system needs to rely on all the data in this collection as its background knowledge while trying to process each story description. Without such a knowledge collection, every time a user inputs a sentence into a system, the system will need to ask countless questions about every simple fact trivial to the human users, and the interaction will be tedious. To give an example, from the story description “His girlfriend is going to the hospital to visit her grandmother”, a human reader can inference a lot of information, such as “He has a girlfriend”, “He is a male”, “his girlfriend is a female”, “He is not married”, “his girlfriend is not married”, “his girlfriend’s grandmother is sick”, “his girlfriend’s grandmother is not died yet”, “there are doctors in the hospital”, “there are nurses in the hospital”, “some of the doctors and nurses in the hospital take care of his girlfriend’s grandmother”....and so on. Some of these inferred sentences might not be true (e.g. her grandmother could be a healthy worker in the hospital), nevertheless it will still be extremely helpful if the system can have a knowledge base that provides all these pieces of information.

The knowledge base used in our Storied Navigation system is the commonsense knowledge corpus derived from the Open Mind Common Sense (OMCS) website [3]. It is a project that aims to collect common sense such as, “You may use an umbrella when it is raining”, “A dog is a common pet”, etc. Currently, OMCS contains over 800,000 English sentences about commonsense, collectively contributed by over 20,000 users from the Web community. Recently, projects collecting commonsense in other languages such as Portuguese, Korean, Japanese, Chinese, etc. are developed based on the approach of OMCS as well [36]. The CYC Knowledge Base [37], is also a huge commonsense corpus, containing more than 3,000,000 hand-entered data entries. CYC differs from OMCS that it uses a formal logic representation to minimize the ambiguity, as opposed to OMCS’ natural language representation.

2.3.2 ConceptNet: A Common Sense Semantic Network

ConceptNet (Figure 1) is an open-source tool for using the commonsense knowledge collected in OMCS, developed by Liu and Singh [10]. It is a semantic network with 20 link types that describe different relations among things, events, characters, etc, and is the major technology used in our fashion system. Example relations in ConceptNet include:

- IsA(A, B) (e.g., “A [dog] is an [animal]”)

- LocationOf(A, B) (e.g., “[Books] are in the [library]”)
- UsedFor(A, B) (e.g., “[Forks] are used for [eating]”)

In the example above, the “IsA” link, connects the two nodes, “dog” and “animal”. Both of these nodes can be in turn connected with other nodes in various links as well. The graphical nature of ConceptNet allows it to perform various inferences by propagating concepts through the connected network, such as affect sensing and theme extraction, as described below.

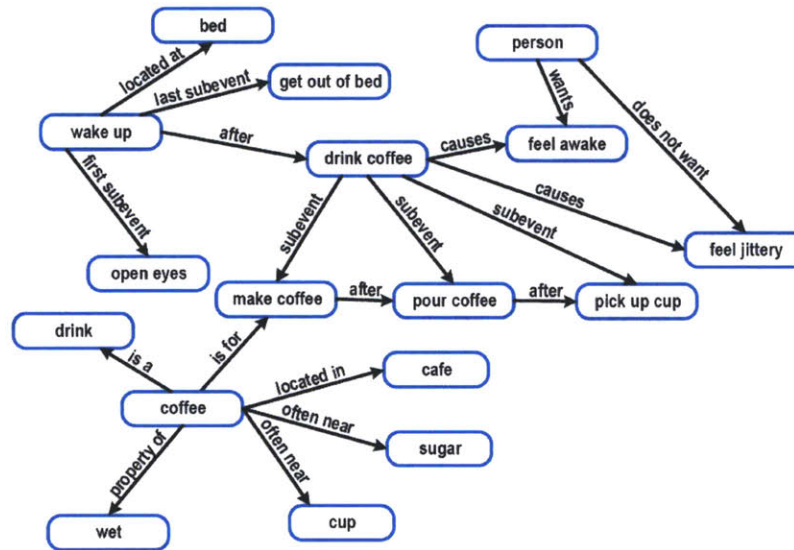


Figure 2-1. ConceptNet [34]

The second reason why commonsense computing technology, particularly ConceptNet, has to be leveraged, is that it can analyze the affect, or emotion, of a piece of story text, even if the text does not contain any terms that explicitly express the emotion (e.g. happy, sad, angry) of the described characters [41]. For example, the story description “Joe’s car crashed in an accident today” carries emotions like surprised, fear, angry, and so forth, while none of these emotional terms appears in the sentence. By performing spreading activation, the system will be able to guess the emotional state of each of the characters in a story description.

2.3.3 WordNet & MontyLingua for Natural Language Processing

WordNet [1] is an electronic lexicon that provides various useful information for English words for computers, including meanings, part-of-speech, synonyms, hypernyms (general categories for a specified word), and so on. It was not designed to be a commonsense toolkit, but its capability of providing synonyms, hypernyms, and word familiarity, is indeed useful in a “commonsensical” way. For example, the verb “inspire” has a synonym “stimulate”, which is a piece of commonsense

information that is not necessarily contained by ConceptNet. As a contrast, ConceptNet can be used to find phrases that are semantically related to the input terms, but not on the word level. For example, the word “eat” can be used to find “breakfast”, “be hungry”, “restaurant”, which WordNet does not respond with [2].

The third reason why commonsense computing technology, particularly ConceptNet and WordNet, has to be leveraged, is that it can relate most input words to other terms that share similar semantics or contexts, such that the semantic similarity between two story descriptions can be evaluated. For example, “John is asking Mary a question about her travel to Europe” can be recognized as a similar story description of “John is wondering about Mary’s trip to Europe.” Indeed, the large area of semantic similarity has a considerable literature to look at, and the approach that we use to evaluate the semantic similarity is relatively simple. However, as the reader will see, our approach is useful enough for the Storied Navigation objective, based on these commonsense computing tools.

Another useful tool is MontyLingua [4], which is a natural language processing tool based on ConceptNet. The Storied Navigation system uses MontyLingua to chop sentences into sub-sentences, to find concise structures for each sub-sentences, to find part-of-speech tags for sentences, and to lemmatize the input words. The detail of using MontyLingua will be introduced in Chapter 3.

2.4 Story Analogy & Story Feature Computation

In the previous sections, we have reviewed Shank’s theory on story indices, derived the story features for videos by analyzing the film “Life. Research.”, and introduced the commonsense computing tools with respective reasons. Nevertheless, there are a few questions that we need to think about, before entering the topic of system design. Namely, a) “Of all the story features derived from the film analysis, what are the possibly extractable ones from free-text story descriptions using commonsense computing tools?”, b) “Similarly, of all the story indices that Schank proposed (theme/topic, goal, result, lesson/observation), what are the possibly extractable ones from free- text story descriptions?”, and c) “For all the collected indices or features, how should we utilize them to achieve the ‘ways to think’ for finding the continued video sequences?” In this section, I attempt to propose answers to these questions.

2.4.1 Extracting Story Features

Again, the story features derived from the film analysis include the content features: 1) the characters, 2) the characters’ interactions, 3) the characters’ emotions, 4) time, 5) location, and the

functional features: 6) whether playing an opening, establishing, ending, or other kinds of functional role in the whole structure, and 7) whether it has to appear at a certain point in the story flow. First, the characters and their interactions can be broken down to, for each sub-sentence in the story description, the subject character, the subject character's action, and the object character of the action. These are information that can be derived relatively easily with natural language processing techniques, since they appear explicitly in the story descriptions. Second, the emotions of the characters can be inferred by performing spreading activation in ConceptNet. Time and location are not necessarily mentioned explicitly in the story descriptions, but, for a story description of a video, these two features can be guessed by referencing other existing videos that share similar characters, actions, and so on. Or, they can be provided by other data sources like GPS or video camera's records as well. For the functional features, both of them require users' input and are not extractable from story descriptions, since, in my perspective, 6) depends highly on the visual characteristics of the videos, whereas 7) depends on the length of the whole story and the main intention of the whole story.

2.4.2 Extracting Schank's Story Indices

If we take a look at the example in section 2.1 again,

- i. "Trying to prove yourself may cause you to do things you don't want to so you can appear cool"
Topic: group acceptance | Goal: appear cool | Actual Results: feel lousy | Lesson: Be yourself
- ii. "Sometimes you act irrationally in a group when you wouldn't if you were alone"
Topic: group acceptance | Goal: appear cool | Actual Results: feel lousy | Lesson: Be yourself

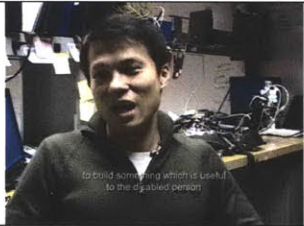

both sentences contain words that are related – on word level – to the topic or theme ("prove" vs. "acceptance"; "group" and "alone" vs. "group"), and they both contain words that are related to the goal ("appear cool" vs. "appear cool"; "act irrationally" vs. "appear cool"). Thus, it seems possible for us to build a system that can select a topic/theme and a goal from respective sets of candidates, by looking for similar terms based on commonsense computing. As for actual results and lessons, they do not share any related terms with the appeared words in the sentences, and they seem to require more deliberations to derive than the previous two. Therefore, for the current Storied Navigation design, we will leave the actual results and the lessons. Also, since it would be difficult for the system to distinguish topic/theme from goal simply by searching for related terms to the words appearing in the story descriptions, we will combine them as a "theme" feature in the system design. The detail of extracting and utilizing this combined theme feature will be introduced in Chapter 3.

2.4.3 Using the Features: Story Analogy

According to the previous two subsections, the list of features that will be extracted from story descriptions by the system has become: 1) The characters, 2) the characters' interactions, 3) the characters' emotions, 4) time, 5) location, and 6) theme, whereas the story structural role of a video will require users' explicit specification. The only question left is, "How are we going to utilize these extracted features to perform the aforementioned 'ways to think' in order to continue users' stories?"

If we look at the eleven different "ways to think" for selecting continued video clips, we will find that the system will only be able to tell whether the last one is true: "it is continued from the previous sequence in terms of time, location, character, event". However, if the system could make analogy between story descriptions using the existing features, it would be possible for the system to tell if d) "it strengthens, extends the story, or elevates the current story to a different level", g) "it presents related meanings to the existing story from a different point of view, context, with a different strategy, or in a different form", and h) "it shares analogous concepts (relationships, attitude, interaction, etc) with or relates conceptually to the previous sequence" are true, since they all require the ability of making analogy between stories.

Table 2-3: "Characters and their Interactions" Features of Two Consecutive Video Scenes

	Characters in the story told	Teacher (Hugh)	Student (Sam)
	Characters' emotions/attitudes	advising, persuasive	Uncertain
	What happened	Hugh convinced Sam to choose the right research direction	
	Characters in this sequence	Teacher (Hugh)	Student (Edward)
	Characters' emotions/attitudes	advising	Uncertain
	What happened	Hugh is helping Edward to make clear his career goal	

To illustrate my theory, let's take h) "it shares analogous concepts (relationships, attitude, interaction, etc) with or relates conceptually to the previous sequence," as an example. It is used during the transition between "Courage in career decisions 2" and "Courage in career decisions 2" in "Life. Research." Table 2-3 shows the comparison between two of their features "the characters" and "the characters' interactions":

As the readers can see, in these two analogous scenes, the characters are either the same or have the same roles, their respective emotions or attitudes are similar, and the subject characters' actions also share similar concepts (e.g. "talking") too. Thus, we can derive this hypothesis: *two story descriptions are analogous to each other if they share 1) analogous character pairs, 2) respective characters' emotions, and 3) the subject characters' actions that share similar concepts*. If this hypothesis holds true, then the system would possibly have four different "ways to think" in terms of story continuation, instead of one.

To see whether the commonsense computing tools allow us to perform such analogy between story descriptions, a few little experiments are conducted, as described in the below paragraphs. Note that these are not formal experiments conducted in any well-designed settings, neither are they used to solve any problems beyond the scope of the problems stated in section 1.2. They are introduced only for the purpose of illustrating how the system design evolved. The formal usability test of the system, which serves as the solution to the stated problems, will be described in Chapter 5.

Table 2-4: Matched Pairs of "Verbs that Share Similar Concepts"

Verb 1	Verb 2	Score
Assist	Help	4.99
Love	Adore	3.03
Assist	Support	2.18
Smash	Hit	1.95
Admire	Like	1.72
Like	Love	0.85
Smash	Run	0.36
Kick	Lie	0.39
Kick	Love	0.13
Assist	Hit	0.01
Assist	Lie	0.01
Adore	Sit	0

To verify the capability of "finding subject characters' actions that share similar concepts", I tried to use ConceptNet to find the matched pairs between 13 randomly-picked verbs, namely, "Admire", "Adore", "Assist", "Help", "Hit", "Kick", "Lie", "Like", "Love", "Run", "Smash", "Support", and "Sit". Briefly speaking, for each of these verbs, the similar concept set is the intersection between the results given by ConceptNet's "get-context" function and WordNet's synonym list. The score of matching two verbs, furthermore, is the sum of the accompanied values of their matched related concepts, derived from ConceptNet's "get-context" function. This method is identical to the

“FindSentenceConceptRepresentation” function in the Storied Navigation system, which will be detailed in Chapter 3. The top 12 matched pairs of these 13 verbs are listed in Table 2-4, from which one can tell that the commonsense computing tools seem to be able to find “subject characters’ actions that share similar concepts” for a piece of story description.

Second, the characters’ emotions can be measured by performing spreading activation in ConceptNet, which has been demonstrated in Liu et al’s paper [41]. Finally, in order to verify commonsense computing tools’ capability of finding analogous character role pairs, the term “analogous” needs to be defined more clearly. In my perspective, a pair of characters is analogous to another pair if the relationships in between the characters of both pairs are similar or identical. i.e. [“teacher”, “student”] is analogous to [“parents”, “children”], because “teacher” and “parents” are both guiding, encouraging, protecting, educating characters to “student” and “children”. Note that the *relationship* is focused on instead of individual identities for performing this analogy [2]. For example, [“teacher”, “student”] and [“parents”, “firemen”] will not be an analogous pair, since the relationships between the characters in both pairs are not similar at all.

Table 2-5: Analogy between Two Characters Based on the Verb (>0 is good enough)

Character1	Character2	Verb 1	Closest Verb 2	Score of this closest verb
Teammate	Friend	Cooperate	Help	6.32
Employer	Boss	Fire	Fire employee	3.61
Family	Friend	Care	Eat	0.11
Designer	Engineer	Create	Build	1.44
Designer	Engineer	Design	Build	1.09
Professor	Teacher	Teach	Teach	5.57
Professor	Teacher	Help	Teach	0.68
Professor	Reporter	Help	N/A	0
Professor	Reporter	Teach	N/A	0

It is not easy, though, for current commonsense computing tools to label “relationships” between arbitrary pair of character roles. Therefore, an easier alternative is to find the analogy between respective character roles across pairs. That is, “teacher” will be mapped with “parents”, and “student” will be matched with “children”. If both [“teacher”, “parents”] and [“student”, “children”] pairs are analogous pairs, then we can assume that the two character role pairs are analogous. Based on this new approach, a character role will be represented by its outward “is capable of” links in ConceptNet, whose end nodes are verbs or verb phrases. Table 2-5 shows the best matched verb end nodes for character pairs. In the first row, for the two characters “teammate”, and “friend”, the best matched verb nodes are, respectively, “cooperate”, and “help”, which has a

score of 6.32, using the aforementioned "FindSentenceConceptRepresentation" metric. "Employer" and "boss", on the other hand, have a score of 3.61. From this table, we can assume that, two character roles are analogous if their matched score in this table exceeds zero. Furthermore, for two pairs of characters [a1, a2], [b1, b2], if the scores of both [a1,b1] and [a2, b2] exceeds zero, than [a1, a2] and [b1, b2] are an analogous character pair.

Based on the two previous experiments, a third experiment is conducted to measure the degree of analogy between concise representations for story descriptions, i.e. [character1's action, character1, character2, character1's emotion, character2's emotion]. The five randomly listed representations are:

- 1) ['help', 'teacher', 'student', 'patient', 'nervous']
- 2) ['teach', 'senior engineer', 'intern', ", 'frustrated']
- 3) ['guide', 'parent', 'child', ", 'frustrated']
- 4) ['hug', 'man', 'woman', 'happy', 'happy']
- 5) ['love', 'husband', 'wife', 'romantic', 'romantic']

Each representation is compared with the other four, and from each comparison a set of five scores is derived, respectively for each entry in the representation. In the design of the list, I suppose the first three representations belong to a same group, whereas the fourth and the fifth representations belong to another. Table 2-6 shows the result:

Table 2-6: Matched Pairs of Concise Representations for Story Descriptions

Target	Best Match(es)	Score
('help', 'teacher', 'student', 'patient', 'nervous')	('guide', 'parent', 'child', ", 'frustrated')	(1.09, 5.95, 209.88, N/A,2.82)
('teach', 'senior engineer', 'intern', ", 'frustrated')	('guide', 'parent', 'child', ", 'frustrated')	(1.01, 0, 5.61, N/A, 5.0)
('guide', 'parent', 'child', ", 'frustrated')	('help', 'teacher', 'student', 'patient', 'nervous')	(1.09, 5.95, 209.88, N/A,2.82)
	('teach', 'senior engineer', 'intern', ", 'frustrated')	(1.01, 0, 5.61, N/A, 5.0)
('hug', 'man', 'woman', 'happy', 'happy')	('love', 'husband', 'wife', 'romantic', 'romantic')	(0.38, 1.20, 17.51,3.60, 3.60)
('love', 'husband', 'wife', 'romantic', 'romantic')	('hug', 'man', 'woman', 'happy', 'happy')	(0.38, 1.20, 17.51, 3.60, 3.60)

As can be seen from the table, the first three representations are grouped together as similar ones, whereas the later two are grouped as another similar set. For the third target representation, i.e. ['guide', 'parent', 'child', ", 'frustrated'], I listed two of the best matched results, because the upper one is scored higher than the lower one for all of the entries besides character 2's emotion. Thus, to choose one of them to be the best match will require weighting functions or parameters for the entries, such that a balanced consideration among these criteria can be created according to the system developers' needs.

2.5 Summary

In this chapter, we introduced Roger Schank's story indices, together with the reason why a storied mind or system needs those indices, the "ways to think" in terms of story continuation based on the analysis of "Life. Research.", the to-date commonsense technology, a combined set of story features based on the commonsense computing technology's capability, and three small experiments that suggest the way commonsense technology can be utilized, i.e. story analogy, in order to perform certain numbers of those "ways to think". Considering the three sub-problems derived in section 1.2, this chapter also suggests possible solutions, as listed in Table 2-7.

Table 2-7: Sum-Problems Stated in Section 1.2 & Proposed Solutions

Problem	Proposed Solution
From a piece of story description, how do we extract features that are potentially useful in humans' "ways of thinking" in story continuation activities?	Using the commonsense technology introduced in section 2.3 to extract the story features: 1) the characters, 2) the characters' interactions, 3) emotions of these characters, 4) time, 5) location, and 6) theme, from a story description.
Using the extracted features, how do we design a metric that measures the distance between two videos' semantic meanings, such that the system can recommend similar alternatives for a specified clip?	Measuring the degree of analogy between the characters, similarity between the characters' emotions, shared related concepts of the verbs, matching the location and time keywords, as well as using the "FindSentenceConcept-Representation" function to measure the similarity between two themes
For a set of candidate videos, how do we design a metric that measures their appropriateness of following a specified one in a temporal order, such that a coherent, inspiring, and intriguing story flow can be produced?	Trying to use commonsense computing tools to perform the "ways to think" for story continuation, derived in section 2.2.

In the next chapter, we will describe how we actually design the Storied Navigation system, based on all these possible solutions.

III System Design & Implementation

"My claim is that storytelling strongly reflects intelligence. Telling a good story at the right time is a hallmark of intelligence" – R. Schank, 1991.

This chapter introduces the interaction design, the algorithms, and the implementation of the Storied Navigation system.

3.1 Interaction Design

The video storytelling activity comprises three major processes, namely 1) to initiate a story by choosing some video clips, 2) to extend the existing story by adding new video clips, and 3) to change the existing story by replacing several video clips. These processes iterate again and again as more and more little stories are made, which can possibly form a big story all together if the storyteller wishes. Therefore, a system that aims to help users tell stories using video needs to provide functions that help them in all three processes.

For the first process, I try to realize the scenario where users can type in story descriptions to initiate their stories, simply because textual narration is the most natural tool human beings use to

Table 3-1: "Ways to think" for selecting continued video sequences, and the required story features

Ways to Think	Characters	Characters' Interactions	Characters' Emotions	Time	Location
Finding a sequence that is continued from the previous one in terms of time, location, character, event	Required	Not Required	Not Required	Required	Required
Finding a sequence that strengthens, extends the story, or elevates the current story to a different level	Required	Required	Required	Not Required	Not Required
Finding a sequences that presents related meanings to the existing story from a different point of view, context, with a different strategy, or in a different form	Required	Required	Required	Not Required	Not Required
Finding a sequence that shares analogous concepts (relationships, attitude, interaction, etc) with or relates conceptually to the previous sequence	Required	Required	Required	Not Required	Not Required

tell stories. I also try to preserve the conventional way of initiating, or making any adjustments to the stories in video storytelling – dragging the clips into the timeline – in the interaction, such that the flexibility is maintained. A search function based on different story features will be provided to facilitate this dragging activity.

The second process is to extend the stories. According to Chapter 2, Table 3-1 shows the “ways to think” a system can potentially perform for selecting continued story sequences, with respect to their required story features. In other words, by analyzing the existing story descriptions, our Storyed Navigation will be able to provide recommendations of “What’s Next?” when the users need it. After the recommendation is given as a list, the user can drag the most desirable one into the timeline. Similarly, the third process, finding video clips to replace existing ones, can be done by finding videos that share the closest story features with the selected ones as well. By providing the two types of recommendations, the system will be able to accompany the users in the process of making series of decisions throughout the activity of building stories, as discussed in the very beginning of section 1.2.

Thus, the storytelling activity will be as illustrated in Figure 3-1. Whenever the users type a story description, ask for recommendations, or search by using some story features, the system will prompts with a set of video clips in respond, which can be used to add into the timeline or to replace existing sequences in the timeline. The three boxes on the left side of this figure, therefore,

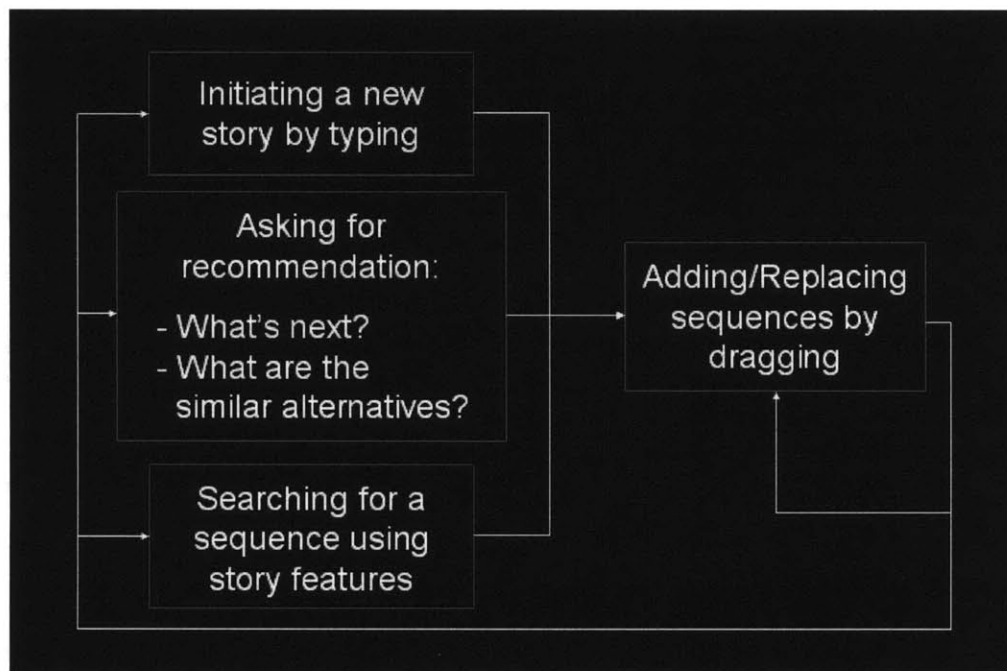


Figure 3-1. Interaction Diagram for the Designed Storytelling Activity

clearly illustrate the functionalities provided by our Storied Navigation, which are also its most distinguishing features compared to other existing video editing/browsing tools.

In the interaction scenario of asking for the system's "What's Next?" recommendations, I chose not to allow the users to select from one of those "ways to think" directly, since they are too complicated and may distract users' focus from the story itself [10]. Instead, a list of simpler criteria are provided, namely, "similar characters", "similar emotions", "similar theme", "following dates", "similar locations", "similar importance level", and "continued story structure". The two criteria that have not been mentioned previously, "importance level" and "story structure" are now introduced. The latter is added because it is an important criterion for two "ways of think", namely "finding sequences that exhibit opening, ending, or other kinds of function to the whole structure", and "finding sequences that establish a new context for a new story part to follow", and is used extensively in my personal experience of video storytelling. "Importance level", on the other hand, is another piece of information assumed to be useful too. Different from the story features, both these types of information are not extractable from story descriptions. They rely on users' additional annotation input while importing the videos.

To get recommendations, the users can either adjust the selection of the aforementioned criteria or use the default one, in which "similar characters", "similar themes", and "continued story structure" are chosen and the rest are not. Similarly, for "Find similar alternatives" recommendations, the criteria will be "by characters", "by emotions", "by theme", "by date", "by location", "by importance level", and "by story structure". As the readers can see, besides date and story structure, all the other criteria for "What's Next?" recommendations are used to find sequences that are similar to the target video clips, because I think such a set of criteria will constitute the four "ways to think" in Table 3-1.

Finally, since the initial sequence-finding as well as both the recommendation functions rely on the story description of the video clips, an annotation interface that helps users input story descriptions easily is also required. Figure 3-2 shows the interaction diagram of the annotation process. In this designed annotation activity, the user will first be asked to select a video clip that he/she wants to annotate, and then to enter a piece of story description for the selected clip. After the system parses the entered story description, there will be two modes for annotating further information, the "basic mode" and the "advanced mode". In the basic mode, users will be asked only to input the roles for the characters that the system recognizes from the input story description. In the advanced mode, on the other hand, the users will be able to modify all the parsed information for each sub-sentence chopped from the story description, as well as the theme, the date, the location, the importance level, and the structural role for this video sequence. The story structural role of a sequence is represented using 8 binary numbers, each standing for one of the following roles:

“begin”, “unfold”, “rise”, “transition”, “conflict”, “negotiation”, “converge”, and “ending”. The importance level, on the other hand, can be either “major” or “subsidiary”. How the system utilizes these pieces of information is introduced in the next section.

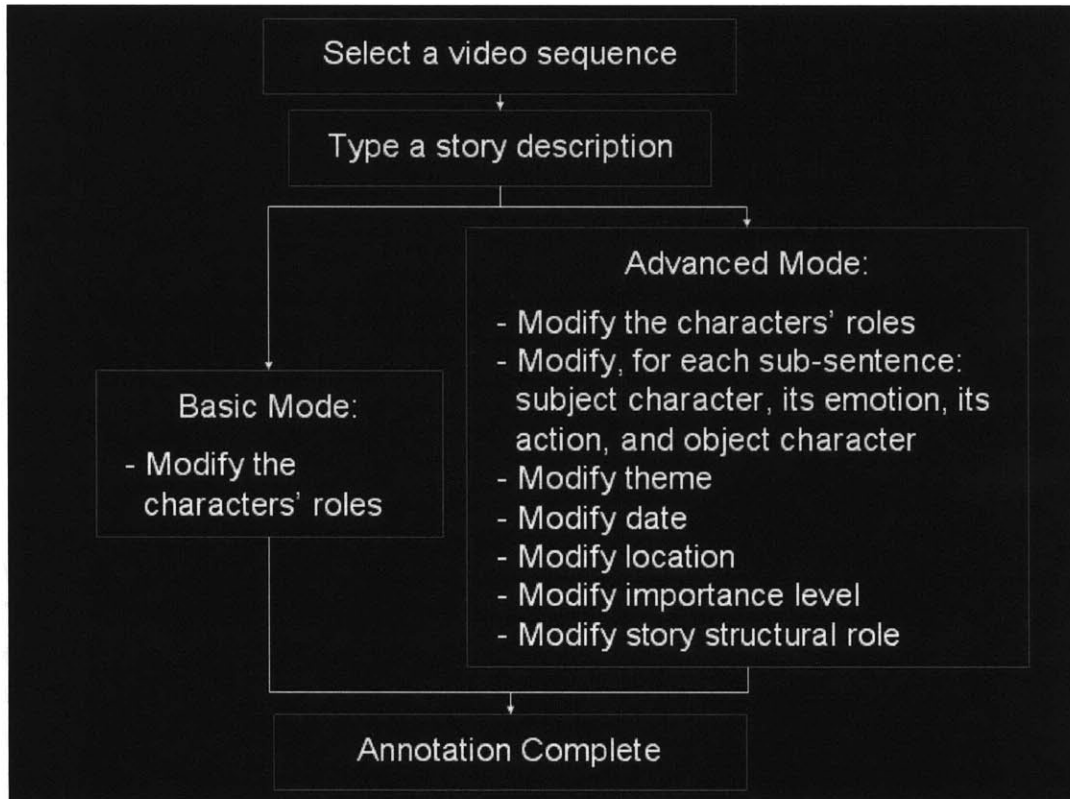


Figure 3-2. Interaction Diagram for the Designed Annotation Activity

Besides the importance level and the story structural role, all the information will be inferred and input to the blanks on the interface automatically after the story description is parsed. Therefore, with the two modes of annotation, both the advanced users and amateurs will be able to use the system in a flexible way.

3.2 Algorithm

This section introduces the algorithms for the three functionalities shown in Figure 3-1, as well as for the annotation interaction shown in Figure 3-2.

3.2.1 Parsing the Story Descriptions

Story description parsing is the most complicated and important function in the Storied Navigation

system, because the storytelling functionalities including initiating stories by typing, “What’s Next?” and “Find similar alternatives” recommendations, and video sequence search based on story features, are all based on this function or its results. The algorithm of parsing each story descriptions is shown in Figure 3-3, and is detailed below.

```

1: procedure StoryDescriptionParsing(TextString t)
2:   sentence_info = InitiateSentenceAnnotation(t)
3:   sub_sentences = ChopIntoSubsentences(t)
4:   primitive_subsentence_representations = new Array
5:   foreach s in sub_sentences do
6:     parsed_result = BasicParsing (s)
7:     primitive_subsentence_representations.append(parsed_result)
8:   foreach p in primitive_subsentence_representations do
9:     for character in (subject of p, object of p) do
10:      gender = RecognizeGender(character)
11:      character = FindNameForPronoun(character)
12:      character = ReplaceOwnershipWithOriginalCharacter (character)
13:      sentence_info.characters.append((character, gender))
14:      sentence_info.subsentence_primitives.append(p)
15:   sentence_info.concept_representation = FindSentenceConceptRepresentation(t)
16: end procedure

```

Figure 3-3. The Pseudo Code of the Story Description Parsing Algorithm

First, the input sentence *t* is chopped into several sub-sentences. The main chopping tool that the “Chop-Into-Sub-sentences” function uses is MontyLingua. However, there are a few cases where it does not perform ideally, including sentences with general conjunctive terms (“because”, “then”, “so”, “therefore”, “if”, “but”, etc.), “wh”-sub-sentences (sentences that contain “who”, “what”, “when”, “where”, “why”, “how”, “which”, etc.), and so on. Therefore, the “Chop-Into-Sub-sentences” function takes care of most of these exceptional cases by adding a period ahead of these terms before sending it to MontyLingua, as well as by duplicating the noun phrase before the “wh”-conjunctive terms when needed. For example, “Ben saw Jane, who was talking on the phone” will be modified as “Ben saw Jane. Jane was talking on the phone” before being sent to MontyLingua. An example of the (input, output) pair of the “Chop-Into-Sub-sentences” function is shown in Figure 3-4.

For each of the sub-sentences chopped by the “Chop-Into-Sub-sentences” function, the “Basic-Parsing” function analyzes it and gives four main outputs, including a subject character, the

subject character's emotion, the subject character's action, and an object character. Besides the subject character's emotion, which is determined by performing spreading activation in ConceptNet, introduced in the next paragraph, the three other outputs are given by the "Extract-Info" function in MontyLingua. The "Basic-Parsing" function also finds other potential human characters (e.g. pronouns such as "her", proper names such as "Tim", role nouns such as "colleagues", etc.) that exist in the sub-sentence. For example, in the sentence "According to Tom's suggestion, we decide to go to that restaurant for dinner", "Tom" is an additional human character that the system should be aware of.

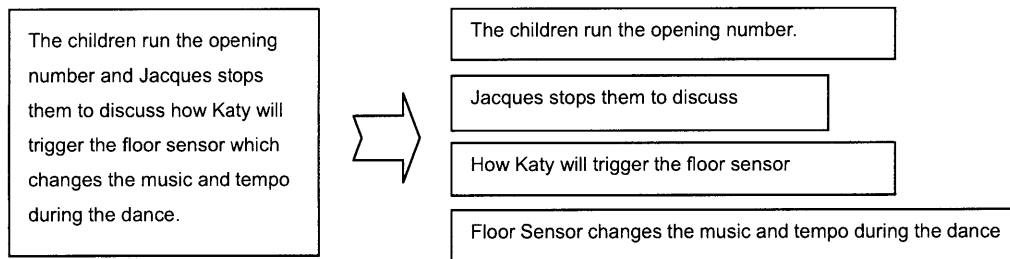


Figure 3-4. Result of the Chop-Into-Sub-sentences Function

The emotion sensing algorithm is based on Liu et al.'s affect sensing algorithm [41]. To describe emotion, or affect in Liu et al's paper, a numeric triple representing three nearly independent affective dimensions is applied. This representation is known as PAD (Pleasure-Displeasure, Arousal-Nonarousal, Dominance-Submissiveness) model [44]. Each of the three dimensions may vary between +10.0 and -10.0. For example, the PAD vector of "love" is [8.72, 6.44, 7.11], and the PAD of "sad" is [1.61, 4.13, 3.45]. Taking the words derived with PAD vectors through human experiments in Mehrabian's paper as "emotion grounds" [44], we can derive the affect value for all the nodes in both the semantic networks by performing spreading activation. To perform spreading activation, briefly speaking, for each of the nodes in ConceptNet with an assigned PAD vector, this PAD vector will propagate outwards to all the neighboring nodes with a decaying weight d ($d = 0.25$ in our system). After all the outward propagations are completed, the new PAD vector for each node in ConceptNet will be derived by averaging all the inward propagated vectors. In the online emotion sensing process for story description parsing, the system will first tokenize chopped the sub-sentence, and find a PAD vector for each of the tokens. For each of the tokens, if the token is not "affect-neutral", i.e., the difference between its PAD vector and [5.0, 5.0, 5.0] exceeds a threshold (1.0 in our system), it will be added into the subject characters' emotional terms. An emotional term can be a noun (e.g. "trouble"), a verb (e.g. "love"), an adjective (e.g. "excited") or an adverb (e.g. "ridiculously"). An example of the "Basic-Parsing" function is shown in Figure 3-5.

In addition, if the sub-sentence is negated, (i.e. it contains terms like “is not”, “doesn’t”, etc), the system will find “counter emotional terms” by finding two emotion ground terms whose PAD vectors are the closest to $[10.0-p, 10.0-a, 10.0-d]$, where $[p,a,d]$ is the averaged PAD vector for the originally found emotional terms in this sub-sentence. For example, for the sentence “Tiffany is not so enthusiastic about Louis’s idea”, the system will find “sad, unhappy” as the new emotional terms by using the original term “enthusiastic”.

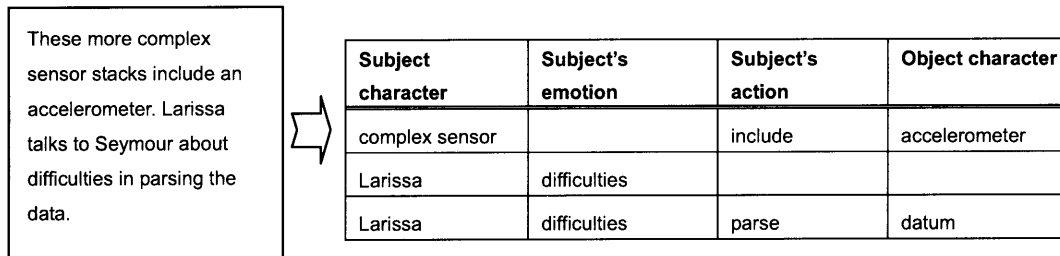


Figure 3-5. Result of the BasicParsing Function

After the “Basic-Parsing” function has parsed all the sub-sentences, three functions are employed to process each of these sub-sentences’ subject and object character, namely, “Recognize-Gender”, “Find-Name-For-Pronoun”, and “Replace-Ownership-With-Original-Character”. First, there are five types of (generalized) “genders” in our system, including “male”, “female”, “group of people”, “uncertain gender”, and “object”. Noun phrases that are not human terms according to WordNet are classified as “object”; noun phrases that are human terms and are group terms (e.g. “team”), plurals (e.g. “friends”), plural pronouns (e.g. “them”) or phrases that contain an “and” and several commas (e.g. “My mother, father, and Joe”) will be classified as “group of people”; noun phrases that are human terms and can be recognized with sexual gender easily according to WordNet (e.g., “Michael”, “her”, “the waitress”) are classified as “male” or “female”; and, finally, noun phrases that cannot be classified into any of the above genders, will be classified as “uncertain gender”, such as proper names that can be used for both sex like “Alex”, or role terms that are suitable for both sexes, like “driver”. Terms with uncertain gender will be assigned with specific genders if mentioned by pronouns in later sub-sentences.

During the processing of each sub-sentence’s characters in the “Find-Name-For-Pronoun” function, what the system does is basically, in one after another previous sub-sentence, looking for an existing subject or object character that shares the same gender and is not a pronoun. If both the subject and object character in a sub-sentence satisfy the result which is searched for, choose the subject character if the pronoun that we are trying to replace is a subject character, and vice versa. For example, for the input sentences “Bob and Katy are my colleagues. They are both talented people.”, there will be two sub-sentences, where “Bob and Katy” and “my colleagues” are the subject and the object character in the first one, while “They” and “people” are in the second

one. Although both “Bob and Katy” and “my colleagues” are desirable for replacing the pronoun “They” in the second sub-sentence, the system will choose “Bob and Katy” because they are both subject characters. In another example, for the sentence “The teacher is angry at Sue because he thinks she should make better progress.”, “The teacher” will not only be used to replace the pronoun “he”, but the system will also assign it with a gender “male”, since “The teacher” appears to be the only suitable “name” for “he”, which in turn confirms that “The teacher” should be a male.

Similarly to the two previous functions, the function, “Replace-Ownership-With-Original” modifies the subject character in the sentence “I met Janet today. Her skirt was gorgeous.”, i.e., “Her skirt”, into a more specific “name”, i.e., “Janet’s skirt”. The reason why these three seemingly simple yet laborious functions are performed, is because the characters’ interactions and emotions heavily rely on their results. Without these functions, the information for performing story analogy would be poor, and most of the “ways to think” in Chapter 2 would not be successfully achieved.

Finally, by using WordNet and ConceptNet, the system constructs a “sentence concept representation” for this sentence by performing the “Find-Sentence-Concept-Representation” function. The algorithm can be divided into the following steps. First, each of the tokens tagged with part-of-speech information, found by MontyLingua, is processed by WordNet to find its synonyms, based on whether it is a noun, verb, adjective, or an adverb. Then, the synonyms are filtered with ConceptNet based on how conceptually related to the token they are, and the remaining ones are considered as consistent with the token’s meaning in the sentence. The conceptual relevance between a token and its synonym is the intersection of their respective spreading activated nodes in ConceptNet [6]. In other words, the more overlapped spreading activated nodes there are, the more the two words are related. We use such a technique to filter out some of the synonyms derived from WordNet, because in most cases there are some synonyms that are less frequently used than others. For example, while finding synonyms for the verb “buy”, WordNet returns [“get, acquire”, “pay”, “get, acquire”, “believe”, “be”] for its five senses. We use ConceptNet to filter out “believe” and “be” as less used, which leads to better results.

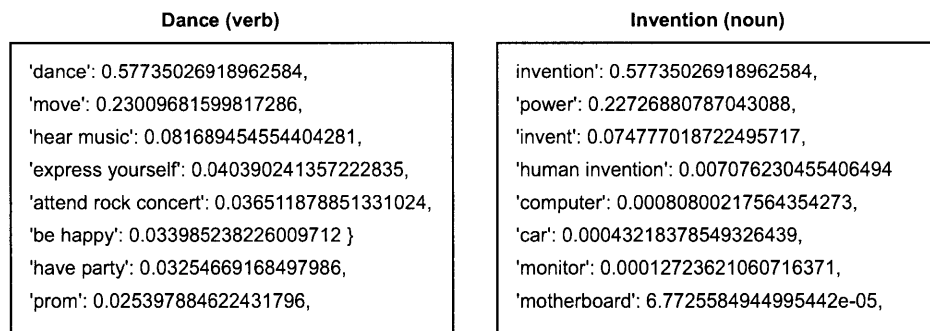


Figure 3-6 (a). Examples of Concept Representations for Single Words

The third step is to find all the related concepts for each token in ConceptNet. The synonyms, together with the token itself, are used as seeds sown in ConceptNet for spreading activation, which results in a concept space that contains a set of nodes and their respective ratings. The

"These more complex sensor stacks include an accelerometer. Larissa talks to Seymour about difficulties in parsing the data."

"Jacques is clear about the contract he will put forward to the children tomorrow morning"

```
'accelerometer': 1.0,
'be complex': 1.0,
'datum': 1.0,
'parse': 1.0,
'sensor': 1.0,
'Seymour': 1.0,
'difficulty': 0.5,
'include': 0.5,
'talk': 0.44721359549995793
'stack': 0.44721359549995793,
'conversation': 0.17812261916432198,
'mouth': 0.17621985483116387,
'have conversation': 0.10473072053771078,
'express yourself': 0.10418710364015524,
'meet friend': 0.080359493280509536,
'meet interesting person':
0.078707153695987589,
'socialize': 0.056677751200669556,
'have party': 0.040747770035070081,
'hang out at bar': 0.027450268913173925,
'tell story': 0.026761090106700872,
'communicate': 0.013516669354515044,
'list': 0.0098090515543927913,
'family': 0.008284428820602539,
'dog': 0.0075287360176080568,
'dinner': 0.0062749847589410792,
regular basis': 0.0049428577046444714,
'several stack': 0.0027781920827814519,
'child': 0.0025580536407328637,
'great difficulty': 0.002493503901871029,
'make bread': 0.0012351533609879673,
'sweat': 0.00087192021508098621,
'retriever': 0.00065271214679100563,
'basis': 0.00059197136863010826,
'spaniel': 0.0004875518937018981,
'parse into interval': 0.0001920561136872541,
'person': 2.2936116420850049e-07,
```

```
'tomorrow': 0.70710678118654746,
'contract': 0.57735026918962584,
'child': 0.5,
'morning': 0.5,
'be forward': 0.44721359549995793,
'put': 0.33333333333333331,
'be': 0.27735009811261457,
'be clear': 0.24253562503633297,
'agreement': 0.21718476723485056,
'set': 0.069848806297102065,
'write': 0.068484223653724655,
'examine thing': 0.046340598852069502,
'procreate': 0.037678319026155933,
'get job': 0.02680925222902799,
'recycling': 0.024024566157186431,
'clean house': 0.021425641787303232,
'set cup on table': 0.016139446074207888,
'help': 0.015652819671396572,
'form': 0.0069470821683251041,
'communicate': 0.0035854698415856939,
'funny something': 0.0034280764807418465,
'give assistance': 0.002788358379936547,
'today': 0.0026083307627548214,
"borrow money from person 's parent":
0.00070775037996839447,
'enjoy play game': 0.00070775037996839447,
'have child': 0.00024474866384917022,
'find mate': 0.00019927147831706586
'not use condom': 9.7618002958615503e-05,
```

Figure 3-6 (b). Examples of Concept Representations for story descriptions

ratings are derived from 1) the relevance scores given by ConceptNet, and 2) the “familiarity” scores, or the statistics of usage frequency, provided by WordNet. The importance of a concept is proportional to the relevance score and the inverse of the familiarity score, since we think less-frequently used words (such as “negotiate”) possess more distinctiveness than often-used ones (such as “go”), in determining similar text fragments. To give an brief example, concepts found for the noun token “party” include “involve”, “wedding”, “pool party”, “party”, “social gathering”, “have brew beverage”, etc.

Note that ConceptNet is aimed to provide “related” concepts, not necessarily “similar” semantics. For example, using the word “purchase”, we can find “store”, “money”, “people”, “credit card”, “cash”, and many other related terms in ConceptNet, but not the word “buy”. Accordingly, it is necessary to find synonyms with WordNet before spreading out into the concepts, which ultimately ensures that the selected video clips can be semantically closer to the input stories. Two example concept representations for single words, i.e. “dance” as a verb and “invention” as a noun, are shown in Figure 3-6 (a), and two other examples for sentences, i.e. “These more complex sensor stacks include an accelerometer. Larissa talks to Seymour about difficulties in parsing the data.” and “Jacques is clear about the contract he will put forward to the children tomorrow morning”, are shown in Figure 3-6 (b).

3.2.2 Annotating Video Sequences

The parsing function described above is used both in the annotation and the storytelling processes. For the annotation activity, the system enters the Basic Mode after the annotation story description is parsed. The user can switch to the Advanced Mode by clicking the “Advanced>>” button, and can complete the annotation process in both modes.

In the Basic Mode, the system shows all the human characters that it finds in the story description, and lists them with their respective roles, as they are critical information for story description matching in the later storytelling activity. More specifically, the system shows the roles for the characters that have been “introduced” in previously video annotations, and leaves them blanks for those that are mentioned for the first time. The users can modify the character-role information as they wish, and the system records modified data into its character-role information. For instance, Figure 3-7 shows the character-role list for the story description annotation “George and Mark’s sister are talking to each other.” “George” and “Mark” are characters that the system knows of, whereas “Mark’s sister” and “George and Mark’s sister” are new to the system.

The advanced annotation mode shows all the information that the users can use to annotate the video sequence, including the character-role information. After parsing the sentence, the system

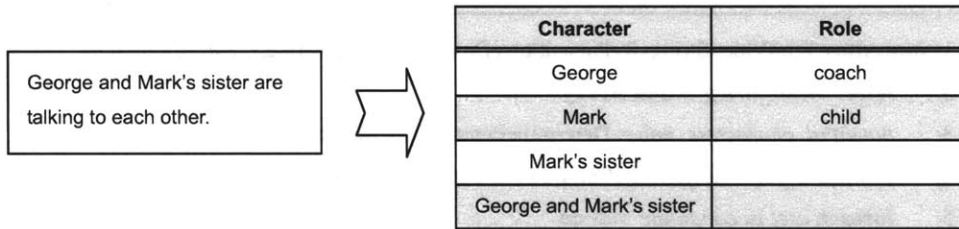


Figure 3-7. The Character-Role List Shown in the Annotation Interface

uses the sentence concept representation and the recognized characters to find the most relevant existing sequences in the same project, and use its theme, location, and date as its default ones. Equation 1 illustrates how I measure the similarity between two concept representations S_{CR} ,

$$S_{CR} = \sum (v_{c,SD_1} \times v_{c,SD_2}) \quad , c \text{ in } CR_1 \text{ and } c \text{ in } CR_2 \quad \text{Eq. 1}$$

where v_{c,SD_i} denotes the value of concept term c in the concept representation CR_i of the i -th story description. It only takes into account concept terms that appear in both concept representations, meaning that these concept terms stand for the intersection of the related concept spaces of the two story descriptions. Using this similarity metric, the system is able to find the most relevant sequences by using a linear combination of S_{CR} and the number of co-existing human characters. After modifying the information, the users can click "Import" to complete the annotation process for this video sequence, and continue to annotate others or start the storytelling activity.

3.2.3 Making Stories by Typing Story Descriptions

To provide users with a set of sequences based on a piece of story description during the storytelling activity, the system executes the "Finding-Video-Strings-by-Parsed-Story-Description" function, illustrated in Figure 3-8

After parsing the user's story description, the system first finds the "potential characters" for the characters described in the story description. That is, if a character appearing in the story description is specific (e.g. a proper name), then the "potential character" for this character will be itself. On the contrary, if the character is not specific, (e.g. roles like "boy" or anonymous human terms like "somebody"), then the "potential characters" for this character will be the union of all the existing characters in the project database that are suitable for this character, according to their roles.

```

1: procedure FindVideoStringsByParsedStoryDescription(TextString t, ParsedResult p)
2:   result_video_string = new Array
3:   potential_characters_info = DeterminePotentialCharacters(p)
4:   candidate_sets = EnumerateCandidateSetsOfCharacters(potential_characters_info)
5:   foreach cset in candidate_sets do
6:     foreach  $\varphi$  in p.primitive_subsentence_representations do
7:        $\varphi$  = ReplaceCharactersWithCandidates(cset)
8:       sequences = FindSequences( $\varphi$ )
9:       sorted_sequences = SortSequences(sequences,  $\varphi$ )
10:      result_video_string = AddBestSequenceToResultString(sorted_sequences)
11:      subsidiary_results = GetSubsidiaryResults(sorted_sequences)
12:     foreach video in result_video_string, subsidiary_results do
13:       video = BuildMatchedStoryDescription(video, t, p)
14:     return result_video_string, subsidiary_results
15: end procedure

```

Figure 3-8. The Pseudo Code for the Find-Video-Strings-by-Parsed-Story-Description Algorithm

Then, the system enumerates all the candidate sets for the characters before searching for the video sequences using these sets. For example, if “Louis”, “Tiffany”, “Henri” are the existing “child” characters in the project and “Jacques”, “Dufftin” are the “teacher” ones, then for the input sentence “the teacher is watching the child improvising his dance,” the system will make the possible set as [“Jacques”, “Louis”], [“Jacques”, “Tiffany”], [“Jacques”, “Henri”], [“Dufftin”, “Louis”], [“Dufftin”, “Tiffany”], [“Dufftin”, “Henri”]. For plural character term in the input sentence (e.g. “children”), the system will include both individual characters and group characters. In other words, if “Louis and Tiffany” is a “children” character in the project and the input is changed as “the teacher is watching the *children* improvising his dance,”, the enumerated list will become: [“Jacques”, “Louis”], [“Jacques”, “Tiffany”], [“Jacques”, “Henri”], [“Dufftin”, “Louis”], [“Dufftin”, “Tiffany”], [“Dufftin”, “Henri”]. [“Jacques”, “Louis and Tiffany”], [“Dufftin”, “Louis and Tiffany”].

After the candidate character sets are enumerated, the system performs the following functions for each sub-sentence, and for each character set: 1) “Replace-Characters-With-Candidates”, 2) “Find-Sequences”, 3) “Sort-Sequences”, 4) “Add-Best-Sequence-To-Result-String”, and 5) “Get-Subsidiary-Results”. First, the system replaces all the characters with the candidates. Then, the system finds annotated video sequences that contain all these characters. If any emotional terms exist in the input story description, the system ignores the sequences that have no emotional terms for the corresponded subject characters.

$$S = \begin{cases} \mu_{CR} S_{CR} + \mu_A S_A, & S_{CR} \geq T_{CR} \wedge S_A \geq T_A \\ 0 & otherwise \end{cases} \quad \text{Eq 2}$$

Third, in the “Sort-Sequences” function, the system calculates the similarity S between the input story description and each of the sequences found in (2). It filters out sequences whose similarity values are below a threshold, and sorts the ones left. The similarity metric is showed as Equation 2, where S_{CR} is the similarity between two concept representations, S_A is the similarity between two PAD vectors, T_{CR} and T_A are the respective lowest thresholds, and μ_{CR} and μ_A are the respective constant parameters.

Equation 2-a shows the similarity between the concept representation of the input story description, and the concept representations of each video sequence in the database. It is basically similar to Equation 1. The only difference is that there are two concept representations for each existing video sequence, namely, the concept representation for the annotation of this video sequence, and that of the theme of this sequence, and they are combined using a linear combination. I choose to use both concept representations in selecting appropriate video sequences because, according to my observation, what the users type tends to be more concise during the online storytelling process, which is easier to find good matches using theme concept representation because using the whole sentence will result in too much noise. On the other hand, not all video sequences will be annotated with a proper theme because of the lack of willingness of the annotator or other reasons, which makes the sentence concept representation still useful.

$$S_{CR} = \sum (\delta_{sentence} v_{c,SD_{input}} \times v_{c,SD_{seq_sentence}} + \delta_{theme} v_{c,SD_{input}} \times v_{c,SD_{seq_theme}})$$

, c in CR_{input} and c in $CR_{sequence}$ Eq 2-a

The affect difference between the input story description and a video sequence to subtract a constant K with the geometric difference D of the two PAD vectors, shown in Equation 2-b.

$$S_A = K - D(A_{input}, A_{sequence}) \quad \text{Eq 2-b}$$

Finally, the system checks whether the best video sequence in the sorted list exist in the result video string or not. If not, the system adds it to the result video string. The rest of the videos found by the algorithm will be collected in to the “subsidiary result list”. Finally, the system finds textual matches between the input story description and that of each of the found videos, in order to show how they are matched on the interface.

3.2.4 Searching Sequences by Story Features

The system provides a functionality that helps user find their materials more easily: Search. They can choose any of the following searching criteria: characters, emotions, theme, location, date, and story structural roles. By, using the aforementioned technique, calculating the number of coexisting characters, the similarity between the emotions in the input and of the video subject characters, the similarity between the concept representations of the input text and of the videos' themes, the number of matches of the locations, matches of the dates, and the number of the matched story structural roles in the selection menu and of the video annotations, the system is capable of finding and sorting videos according to what the user is looking for.

3.2.5 Recommendations: "Find Similar Alternative" & "What's Next?"

The system performs the two types of recommendations, actually performing a set of searches based on the selected criteria, using the information of the selected video as input. For finding similar alternatives, the system takes the characters of the selected video. To perform character search, it takes its themes to perform theme search, takes its story structural roles to perform story structure search, and so on. In this case, the emotion search is incorporated into character search, since the ideal result is not a set of videos whose identical or similar emotional terms are used to describe different subject characters.

On the other hand, the algorithm of performing "What's Next?" recommendation is similar to finding similar alternatives, except for the two criteria, "date" and "story structural role". If the readers refer to the interface shown in Figure 4-12, the option labels for these two criteria are "following dates" and "continued story structure", in addition to the others that are similar to those on the find-alternatives interface: "similar characters", "similar emotions", "similar theme", "similar locations", and "similar importance level". This is because, I think the characters, emotions, theme, locations, and importance level of a following sequence need to be similar or identical to the previous one, in order to perform the four concluded "ways to think" in Table 3-1. To look for video sequences that have "following dates", the system simply finds sequences whose dates are after the selected one. Whereas for finding "continued story structure", the system performs story structure search by using a new 8-ary vector, each of whose binary numbers shifted one-step forward from the selected video's story structural role vector. For example, if the selected video has a story structural role (1,0,0,0,1,0,0,0), which stands for "begin", "conflict" and none of the others, the system will perform story structure search using the new vector (0,1,0,0,0,1,0,0), or ("unfold", "resolution").

3.3 Implementation

The system is implemented in two parts: the interface part and the back-end algorithm part. The algorithm is written in python, and the interface is written in Adobe Flex. The two parts communicate with each other through a HTTP port, and thus the python code is run as a script for a website. The system can be viewed with most of the modern web browsers like FireFox as long as Flash 9 is installed. The reason why I chose python is that it is the native language of MontyLingua and ConceptNet, it is platform-independent, and it is very easy to use. The reason why Flex is chosen, on the other hand, is that it is cross-platform, it supports functions like drag-n-drop in web pages, i.e., it supports Rich-Internet-Applications (RIA), and it provides useful tool kits like list control, date selection control. The implementation took about four months, three for implementing and debugging the python code, and one for building the Flex interface, and was completed in early July 2007.

3.4 Summary

This chapter introduces how the Storied Navigation system was designed and implemented. It describes the algorithms that the system uses to facilitate the annotation interaction, and the storytelling activity including initiating by typing story descriptions, asking for “Find Similar Alternatives” or “What’s Next?” recommendations, as well as sequence search by story features. In the following section, the readers can get a clearer view of how these functionalities are executed and visualized in the interface.

IV Using the Storied Navigation System

"So the issue with respect to stories is this: We know them, find them, record them, manipulate them, use them to understand the world and to operate in the world, adapt them to new purposes, tell them in new ways, and we invent them." – R. Schank, 1991.

4.1 Video Annotation

There is no "story world" to navigate if there is no story in the corpus. Therefore, importing videos into the system is the first step toward enjoyable Storied Navigation activities. When a user opens the video importing interface in the Storied Navigation system, he/she will see a list of videos that are not imported (or annotated) yet. Clicking the videos in the list will result in the playback in the left video player, as shown in Figure 4-1 (a).

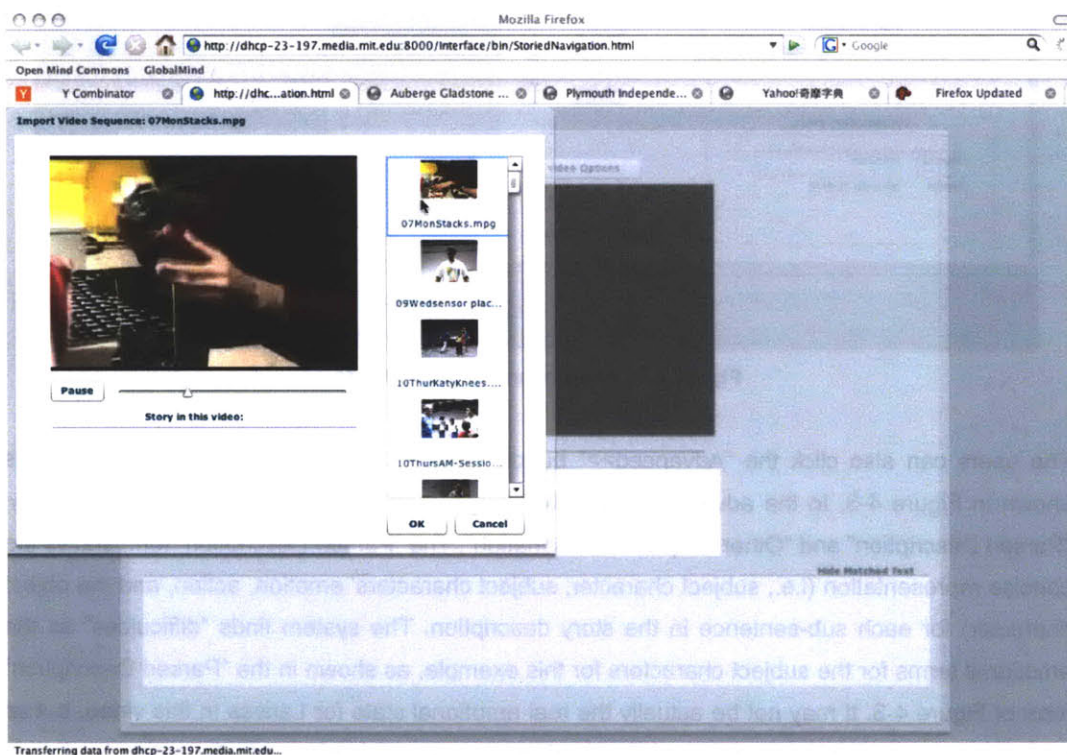


Figure 4-1: Annotation (a)

The user can input a story description in the text area below the video player. In this example, the

user input “These more complex sensor stacks include an accelerometer. Larissa talks to Seymour about difficulties in parsing the data.” After clicking the “OK” button, the system will parse users’ story description, and enters the basic mode, as shown in Figure 4-2.

In the basic mode, only the character-role information is shown to the users. Figure 4-2 shows the two detected characters, “Seymour” and “Larissa”, with their respective roles, “professor” and “graduate student”. This is because, both of their roles are already known by the system, i.e. both have appeared in the previously-imported video sequences. If there is any character whose role is unknown, the system will also list the character, and leave its role blank.

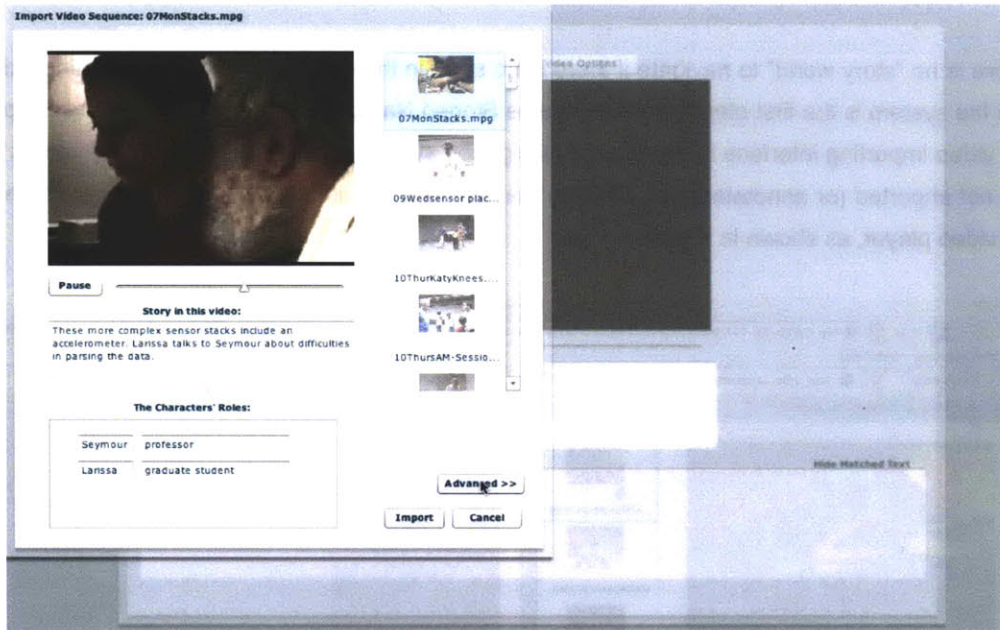


Figure 4-2: Annotation (b): Basic Mode

The users can also click the “Advanced>>” button to enter the advance annotation mode, as shown in Figure 4-3. In the advanced mode, two major forms will appear on the right, namely, “Parsed Description” and “Other Sequence Information”. The “Parsed Description” form shows the concise representation (i.e., subject character, subject characters’ emotion, action, and the object character) for each sub-sentence in the story description. The system finds “difficulties” as the emotional terms for the subject characters for this example, as shown in the “Parsed Description” form of Figure 4-3. It may not be actually the real emotional state for Larissa in this video, but as far as the emotional characteristics “difficulties” can express, this piece of information is already much more useful, than if the user inputs nothing.

On the other hand, the “Other Sequence Information” form shows the other types of information for

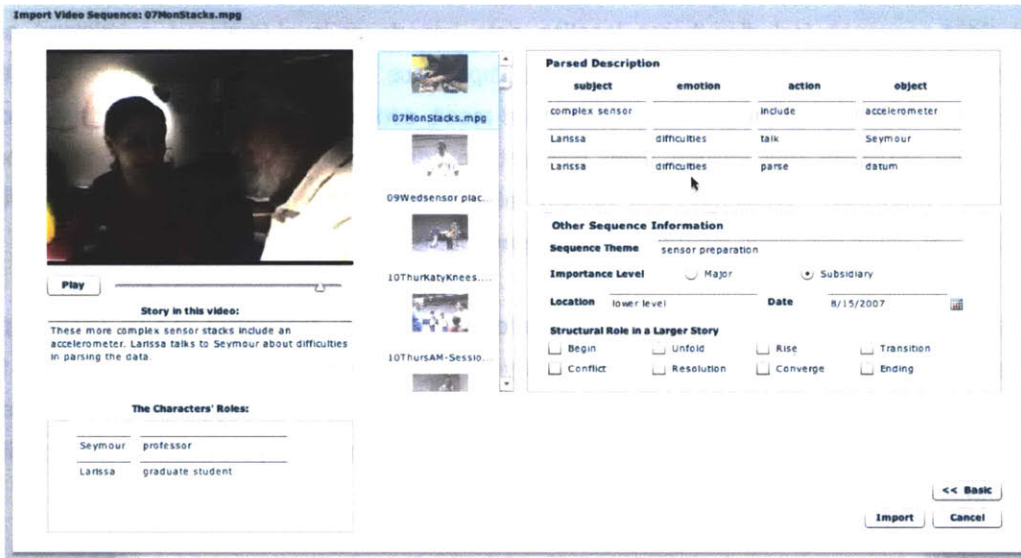
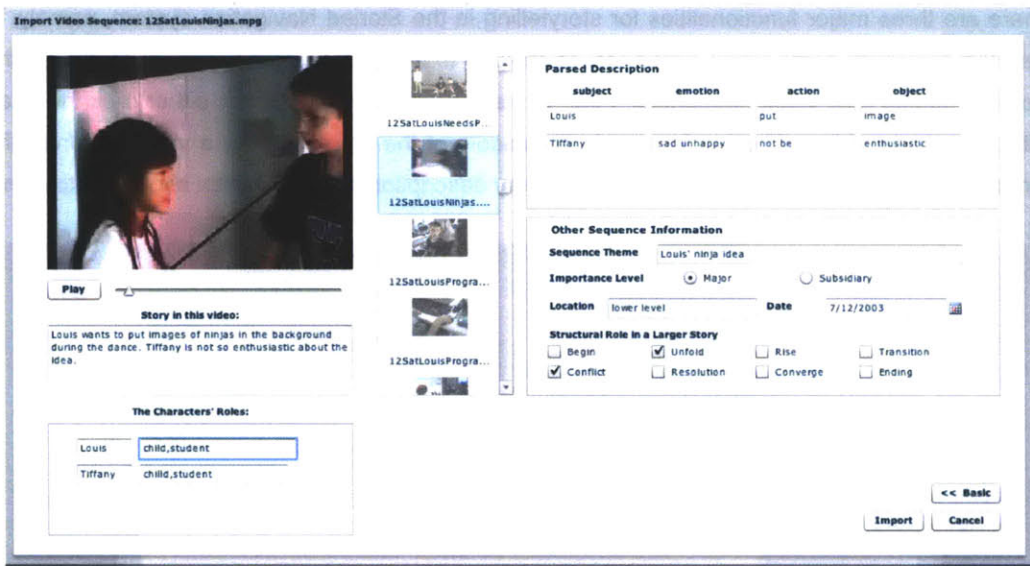


Figure 4-3: Annotation (c): Advanced Mode



Parsed Description

subject	emotion	action	object
Louis		put	image
Tiffany	sad unhappy	not be	enthusiastic

Figure 4-4: Annotation (d): Emotion Detection for Story Description with Negation

this video sequence, including theme, location, date, importance level (major vs. subsidiary), and structural role in a larger story. Aside from the importance level and the structural role, all the information is shown as the parsed or inferred result, according to the input story description. The story structural role include eight different options: "begin", "unfold", "rise", "transition", "conflict", "resolution", "converge", and "ending", all originated from my personal experience of film editing. The users can leave everything unchanged, or modify things as they wish.

Figure 4-4 shows the advanced mode interface for another story description, "Louis wants to put ninjas in the background during the dance. Tiffany is not so enthusiastic about the idea." Again, for the emotion of the subject character "Tiffany", the system found "sad, unhappy", even though there is no such terms in the story description. This example shows that the system has some capability of inferencing for story descriptions that have negation.

4.2 Storytelling

There are three major functionalities for storytelling in the Storied Navigation system, namely, to start the story with some video sequence, to replace existing video sequences, or to extend the story by adding new video sequences. First, to start the story, the user can either click the "start with something arbitrary" button located in the middle of the timeline, drag a video clip from the "annotated video" list to the timeline, or type a story description in the text area above the timeline.

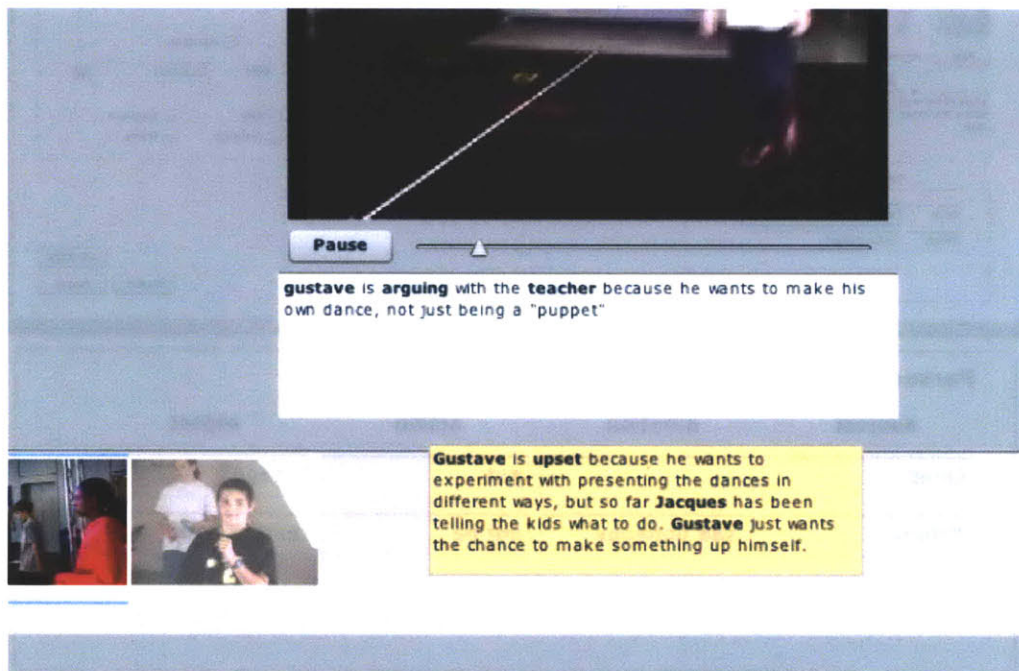


Figure 4-5: The Result for a Typed Story Description (a)

Figure 4-5 shows the result of parsing a piece of story description, "gustave is arguing with the teacher because he wants to make his own dance, not just being a 'puppet'." The system responds with two video clips, which together make the most suitable video stream in response to the story description according to the algorithm. That is, the two videos is a set of result. If the user wants more sets of results, he/she can click on the video clips and choose "show more searched results" from the menu. The video player automatically plays the returned video clips from the first one, and the popup yellow box shows the annotated story description for the video clip that the system is playing back or the mouse cursor points to. In both the input text area and the popped-up boxes, the matched characters, actions, and emotional terms will be highlighted in bold face, such that the users can tell how the clips are selected. In Figure 4-5, "gustave" is matched with "Gustave", "arguing" is matched with "upset", and "teacher" is matched with "Jacques", which are all correct matches. In another example, shown in Figure 4-6, the matches between the input text and the popped-up text are ("gustave", "Gustave"), ("teacher", "Jacques"), ("make", "making"), and ("own", "own"), which is correct too.

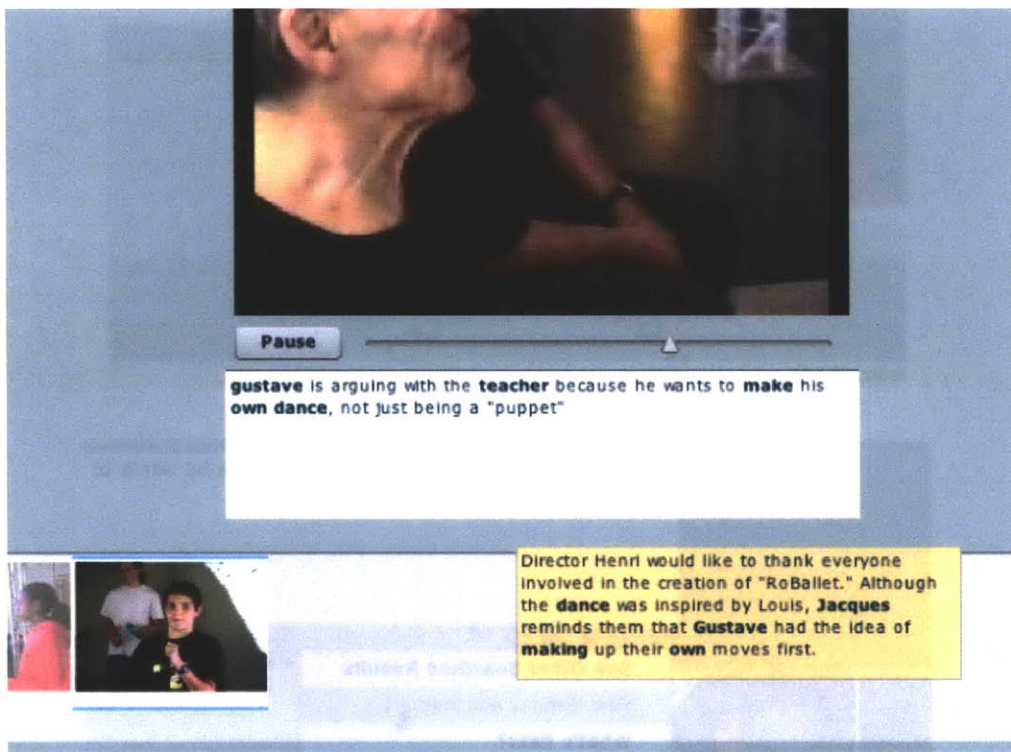


Figure 4-6: The Result for a Typed Story Description (b)

The input story description is not necessarily very specific. That is to say, if a user is not familiar with the video corpus, he/she can input sentence like, "someone is happy", or "people are discussing about something", and the system will still try to prompt with the best result that it can

find. Therefore, a user can gradually “enter” this story world, even if at first he/she did not know much about this word.

After the story is initiated, the user can remove video clips that they do not like by clicking “remove from timeline” in the popped-up menu (Figure 4-7), or they can replace the existing video clips with recommended alternatives (Figure 4-8). A menu will pop up for users to select the criteria for finding alternatives after the user click the “Find Similar Alternatives” menu option, and seven criteria will be listed, including “by Characters”, “by Emotion”, “by Theme”, “by Date”, “by Location”, “by Importance Level”, and “by Story Structure”, as shown in Figure 4-9. In the default state, all of the criteria are used.

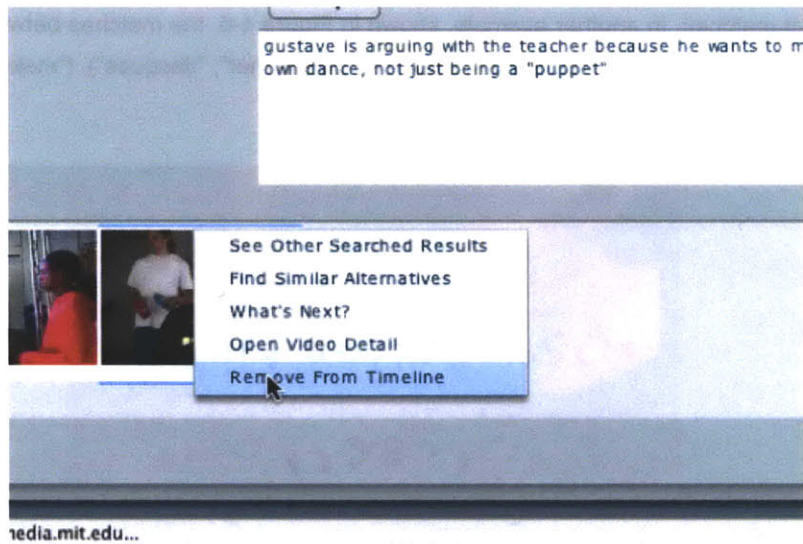


Figure 4-7: Removing a Video from the Timeline

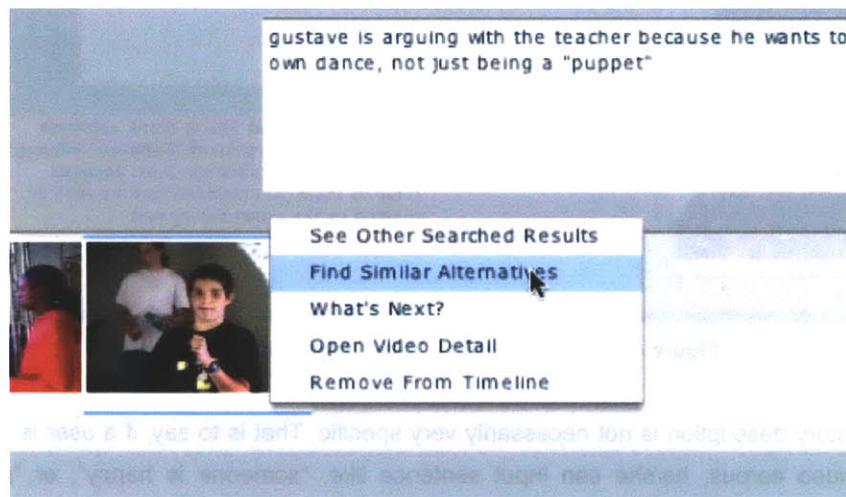


Figure 4-8: Finding Alternatives for an Existing Video (a)

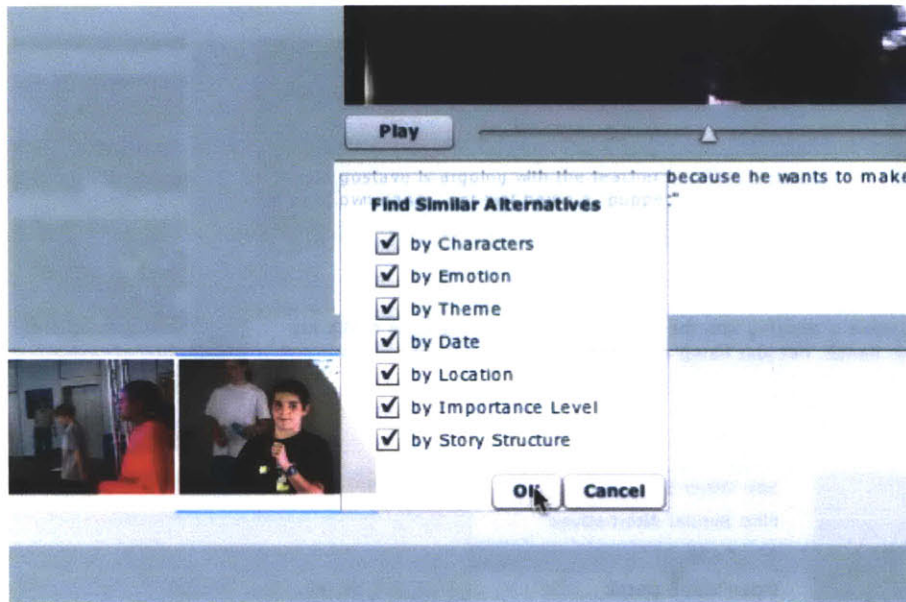


Figure 4-9: Finding Alternatives for an Existing Video (b)

After clicking “OK”, the system will prompt with a list of videos that it finds similar to the target video clip, according to the selected criteria. And, when the user clicks on any of the listed alternative videos, a menu will pop up, and the user can either choose to see the detail of this video clip by opening the annotation interface in advanced mode, or he/she can select “replace with this video” to replace the target video in the timeline with this one, as shown in Figure 4-10. Thus, the process of finding alternatives for an undesirable clip is finished.



Figure 4-10: Finding Alternatives for an Existing Video (c)

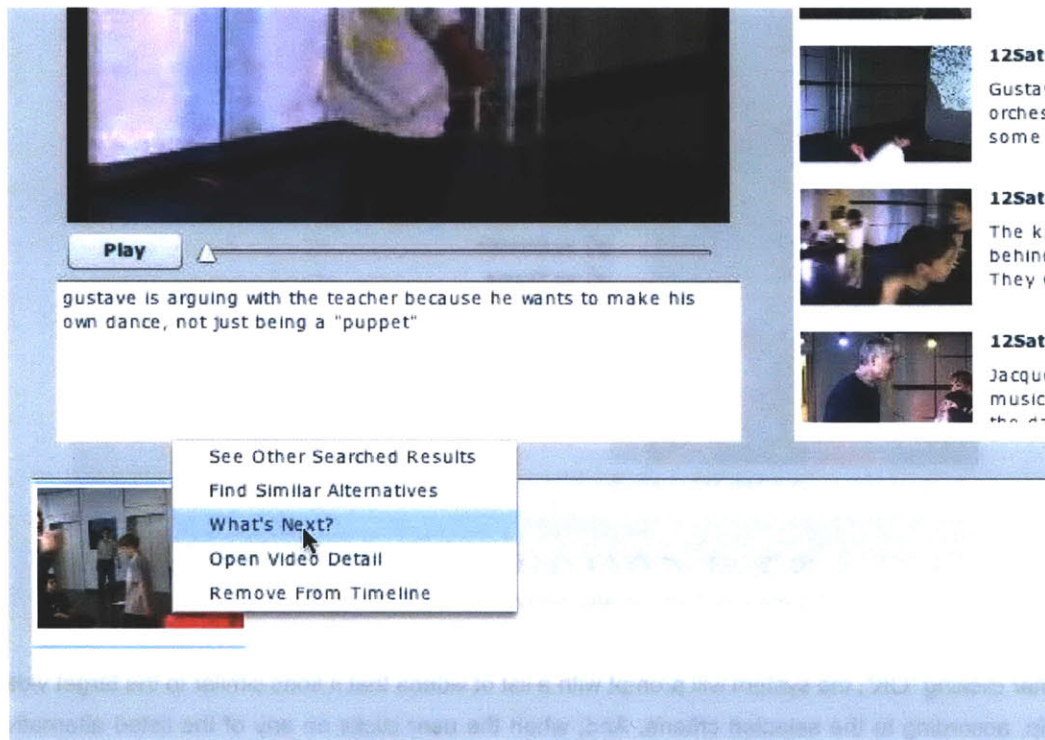


Figure 4-11: Finding “What’s Next?” for an Existing Video (a)

Finally, the stories need to grow. There are a number of ways of extending the existing story, including dragging video clips from the annotated video list, typing more story descriptions into the text area, or using the “What’s Next?” recommendation function of the system (Figure 4-11). Similarly to finding alternatives, there are also a number of criteria to choose before executing the function, including “similar characters”, “similar emotions”, “similar theme”, “following dates”, “similar locations”, “similar importance level”, and “continued story structure” (Figure 4-12). In other words, besides date and story structure, the system will use the criteria to search for video clips in the same way as finding similar alternatives, as stated in Chapter 3.

Suppose the user chooses “similar emotions” as the only criterion, as shown in Figure 4-12, the best searched result, shown in Figure 4-13, would be “Louis wants to put ninjas in the background during the dance. Tiffany is not so enthusiastic about the idea.”, which does share similar emotions (“upset”: “not enthusiastic”). However, if the user select “similar theme” as the only criterion, then the best search result would become “Glorianna asks Jacques about how he plans to involve the children in making decisions about different aspects of the performance.” (Figure 4-14), which is, surprisingly, a nice transition from the previous clip “Gustave is upset because he wants to experiment with presenting the dances in different ways, but so far Jacques has been telling the

kids what to do. Gustave just wants the chance to make something up himself.” Standing from a filmmaker’s perspective, I think this is a very interesting, and potentially useful cut.

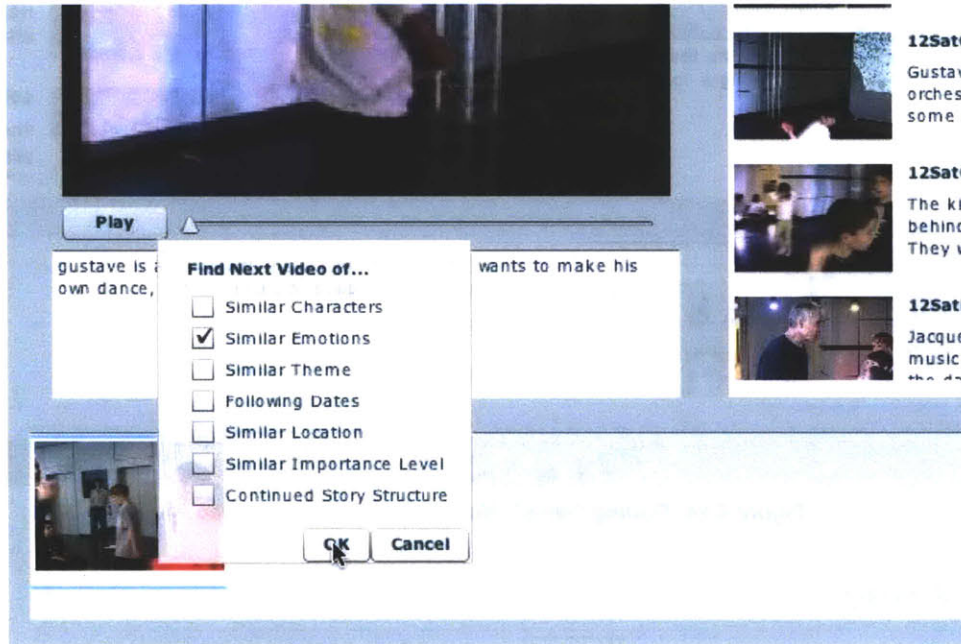


Figure 4-12: Finding “What’s Next?” for an Existing Video (b)

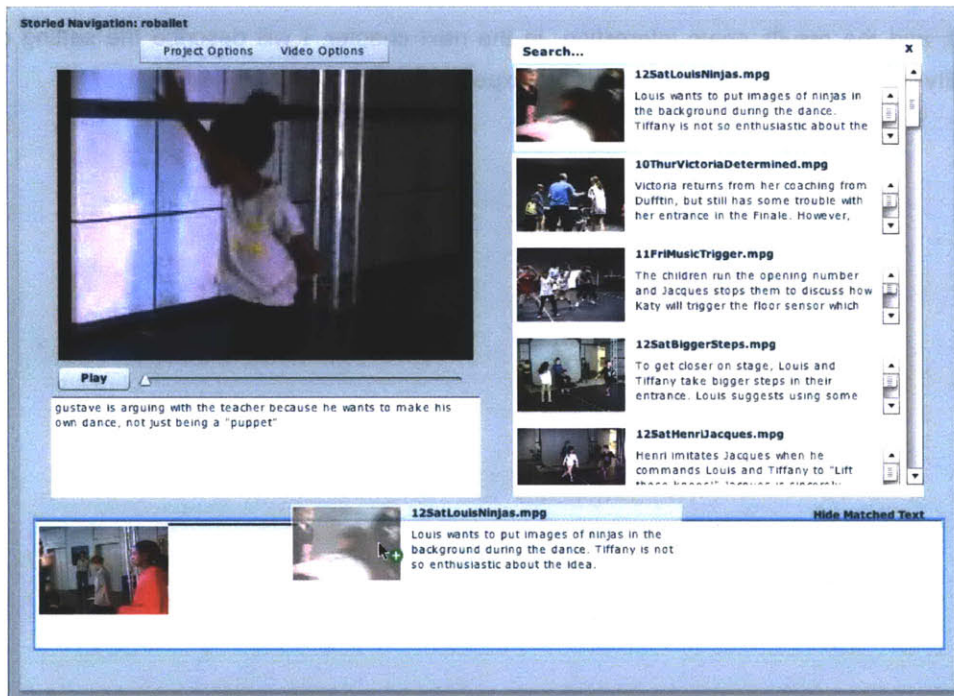


Figure 4-13: Finding “What’s Next?” for an Existing Video (c)

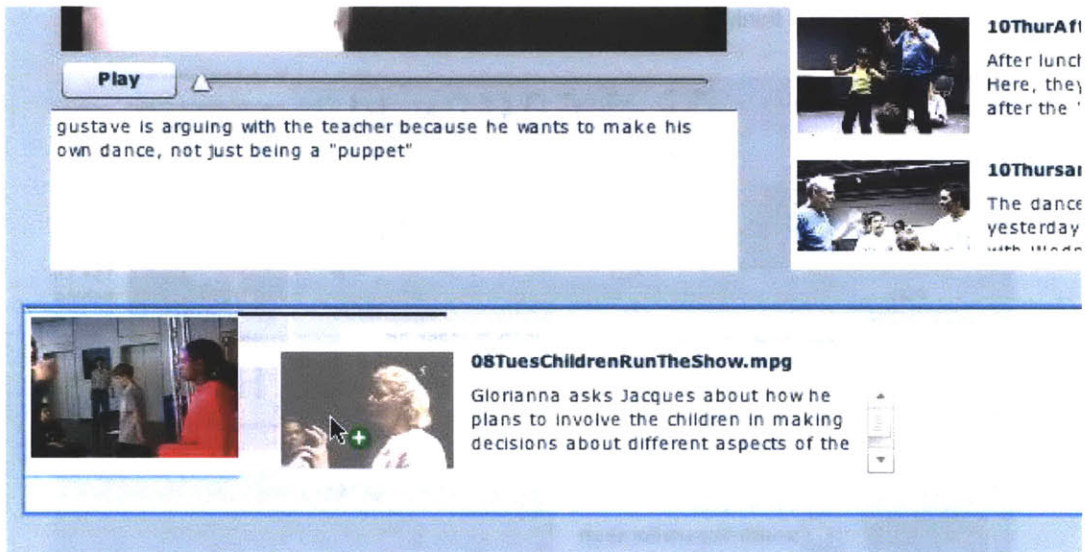


Figure 4-14: Finding “What’s Next?” for an Existing Video (d)

4.3 Summary

In this chapter, the usages of two major activities, namely, video annotation and storytelling, are illustrated. As the readers can see, the functionalities match what I have described in chapter 2 and 3, and the results seem interesting. In the next chapter, I will describe the setting of the usability experiment, and the analysis of the experiment result.

V Observation & Evaluation

“...I can imagine doing that with my videos quickly, if I knew I had your software on my laptop...I will sort of download it, with a quick, quick annotation, and if I don't have time to make my movies now, which happens a lot, and I want to do it, let's say a month from now...It's very usable. Exactly the way it is.” – A user of Storied Navigation.

This chapter is a report of how the users use the Storied Navigation system. There are two major user studies; one aims to observe how users make video stories from a corpus that they are familiar with using our system, and the other one aims to observe how they browse an unknown video corpus. Both studies were conducted in a qualitative way, as detailed in the latter sections, and each subject was asked to fill a questionnaire on the 5-point Likert scale, after the study was completed. There are seven subjects in the first study and two in the second one, because there are not too many people who are familiar with our corpus and accessible at this point.

The video corpus that I use is the Roballet corpus, which is a video collection of a two-week workshop held in summer, 2003. The goal of this workshop was to investigate how technology may be integrated in the creation process of art, as well as new ways of learning. The location of the workshop was the MIT Media Laboratory, and the staff was composed of the professors in the laboratory, the graduate students, and a few professional dancers from outside the lab. There are over 100 video sequences in the database, each ranging from tens of seconds to about three minutes, and 81 out of these sequences were annotated before the studies took place.

In the following sections, I will talk about how I conducted the studies as well as their respective observations and findings, the analysis of the questionnaires, and a summary. While I will mention some results of the questionnaires in Study 1, a more complete analysis of the questionnaires for the two studies will be discussed together after both of their respective analyses are introduced.

5.1 Study 1: Using the System to Make Video Stories

5.1.1 Study 1: Design & Setting

The goal of this study is to investigate whether the design of the Storied Navigation system serves as a solution to one of the problems proposed in Chapter 1: “How can a system help users to develop their story threads?”. Both of the subjects are participants in the Roballet workshop, and

have experience with video editing. The procedure of the study goes as follows. First, the subjects were asked to provide their general information, including what kind of video editing system they usually use and how many hours they have ever spent on video editing. Then they go through an introduction to the system's functionalities, including "edit-by-typing", two types of "edit-by-recommendation", i.e. "What's Next" and "Find Similar Alternatives", the search functions with various criteria such as "by characters", "by themes", "by story structural role", etc, and the annotation. Each of the instructions contains a few examples that they are asked to follow, in order to become familiar with the functions. After the instruction, which typically takes about 30-45 minutes, the subjects are asked to do two warm-up examples for the real storytelling task. The example stories that they are asked to follow and make stories are:

Example 1:

Henri is one of the children in Roballet. He became directive toward other children in the workshop. Please make a story about Henri in which he 1) helps other people make animation, 2) decides what other kids should do during the rehearsal, and 3) sits in the audience to see how the other children do during the rehearsal.

Example 2:

Running a workshop often requires one after another discussion between the staffs. Please find four sequences, each of which containing some discussion participated by some of (but not limited to) these characters: Jacques, Roger, Seymour, Glorianna, David, and Duffin. Make the story flow most desirable according to your own sense of "good story".

They are free to talk to me or ask questions about the system during or after they complete the examples, and can skip the second example if they want to jump to the real tasks. They can also only do one story in the task if they feel tired or because of any other reasons. The two instructions of the studies are:

Story A:

Roballet is a workshop where the researchers try to investigate new ways of *learning*. That is, learning can take place by creating things, by emulating others, by collaborating, by problem solving, by teaching, by planning, or even by doing the opposite way of what's being told. Please use the system and all the materials that it has, and make a story by following the topic "*Learning* in Roballet". You can start by focusing on particular aspect of learning if you like, or you can start by focusing on certain particular characters too. The length of this story is not strictly assigned, but the suggest length is between 10-20 sequences. Please assume the result can be

fine-tuned by using any commercial systems like Final Cut or Premiere, and the current task is only to develop your most desirable story line. Also assume that your audience will be the viewers on the Internet who wish to learn more about new possibilities of learning activities, as well as the Roballet workshop, by watching your video story.

Story B:

Roballet is also a workshop where a new form of art is created by introducing technology into the creating process. Please, similarly to the previous task, make a story that is related the theme "*Technology vs. Art* in Roballet". You can create your own topic based on your own perspective, such as "The useless technology" or "The next, exhilarating era of art". Again, please assume that the result can be fine-tuned by using any commercial systems like Final Cut or Premiere, and the current task is only to develop your most desirable story line. Also assume that your audience will be the viewers on the Internet who wish to learn more about combining technology and art. The suggested length of the story is, again, between 10-20 sequences.

The difference between the descriptions of the examples and the real tasks is, the latter gives more general, broad topic, whereas the former is more specific and allows less freedom. The difference between the two stories in the real task, furthermore, is that Story B gives a hint of controversy, while Story A does not.

<p>What's the story? Tell me about the stories that you made. What was the first story about? How did the story flow? How did it begin? How did it end? Does it have a climax? What was it? Who are the main characters in this story? What aspect of learning did you focus on?</p> <p>Developing the story How did you decide the characters of this story? How did you find the focus of this story? What did you think about first? About when did you decide what this story is about? What do you think was the reason that brought you there? How did you develop the story? What are the functions that you use the most and why?</p> <p>Functions What were the most useful functionalities in the process? Why? By which criterion do you find the search function most useful?</p> <p>Going back to the annotation Did you edit the annotation during the storytelling process? About how many? Under what kinds of circumstances?</p> <p>Difference from the past experience What is the major difference between this storytelling experience from your experiences in the past?</p>
--

Figure 5-1: Sampled Questions in the Interview of Study 1

After the completion of the tasks, despite whether or not they complete both of the stories, an interview of what is the story about and how they feel about the system is conducted. Example questions that I used are listed in Figure 5-1.

5.1.2 Study 1: Observation & Results

There are two subjects in the first study, the storytelling tasks, referred as Subject 1 and Subject 2. The detailed finding from their storytelling tasks are discussed separately in the below subsections. On the other hand, according to the questionnaire on the 5-point Likert scale filled by the two subjects in this study, both subjects “strongly agree” that they can use the system to edit videos, both find using the system “very enjoyable”. They both find the “edit-by-typing” function “very useful”, and both think the “Find Similar Alternatives”, “search by character”, “search by story structural role”, and “search by keyword” “useful” or “very useful”. They both think that the interface is “intuitive” or “very intuitive”, and none of them finds any function “not useful” or “not useful at all”.

5.1.2.1 Study 1: Subject 1

Subject 1 was one of the staff of Roballet who participated in planning the logistics and schedule of the whole workshop, and has 50-100 hours of experience of video editing. Her most familiar editing system is iMovie. She likes the children in the workshop very much as well as the video corpus. She spent about 40 minutes finishing the instruction, and about 75 minutes finishing Story 1. She did not proceed to make Story 2, but we spent much time discussing about her questions and thoughts about the system.

Briefly speaking, Subject 1’s learning story was made by first typing one sentence using the “edit-by-typing” function. Then, by asking for “What’s Next?” based on the first result only, she found almost all the sequences that she needed for completing the whole story. The sentence that she typed initially was “Learning in Roballet, exploring Body language along with mathematical language”, and the retrieved sequence was a sequence of a meeting after the workshop was ended, where the staffs were talking about their own perspectives of how dance and technology should be emphasized without detouring from the main goal of the workshop. Here are two passages quoting from Subject 1 with regard to this result:

- “This is right on the topic that I ask...so this is a perfect clip”
- “...from the first one I derived uh, what’s next. But the first one is so rich. It actually has every single component I put in my sentence. That the ‘What’s Next’ that goes back to explain it...each of the little segments is in there. So I think it works nicely”

During the interview, Subject 1 gave me an incredible number of interesting observations about Storied Navigation, which I categorize as:

- i. Improving Efficiency for Documentary Video Editing
- ii. Quick Annotation. New Way of Organizing the Data
- iii. Annotation Interface Helps Understanding of the Logic Behind
- iv. Paths as Materials
- v. Theme for the Clips and also for the Paths
- vi. Through the Different Corpuses; Through the Different Times
- vii. Flexibility and Encouragement of Collaboration

Below I discuss about each of these points respectively.

- i. Improving Efficiency for Documentary Video Editing

She pointed out that it would be much more difficult to use editing software like iMovie to build such a story, because the time spent for understanding is much longer. Upon the question “What if you just use iMovie from scratch without it?”, Subject 1 reacted, “Oh, that would be incredibly hard. Because it’s one week of workshop. No, what am I saying, a two-week workshop. And, it was intense. I remember we were here at eight o’clock or seven in the morning, and about 8pm...easy. Everyday. and they’re filming constantly....we didn’t have enough, you know, people filming, but, in a way that’s good (laugh). Because you have two people filming full time, I mean, that’s a lot of films. So now if I want to go with this topic, that means I have to watch, how many hours?... A lot.” She indicated that it is much easier to find the things that she wants using Storied Navigation, which helps decreasing the time that she needs to spend for editing a story using documentary videos.

- ii. Quick Annotation. New Way of Organizing the Data

In terms of annotating the videos, she expressed as if a different mindset of organizing video data has emerged from the design of the annotation interface as well as the overall usage of the system, “...it’s hard to go back to something that has been done. But as you produce your material, and you know you can have it applied to a software like this, while you’re doing it you prepare it. You can chop it. So that at the end of the day you come, you just fit it to your software and you make your annotation quickly, and it is ready to use later. It is different to, you know, going back to whole box of films and annotating them, I think.” It raises an interesting question worthy of further investigation: “Does Storied Navigation affect the way people view annotation and manage their media files”, and I would like to see if we can conduct a long-term experiment in the future that tries to answer this question in particular.

iii. Annotation interface helps Understanding the Logic Behind

Subject 1 also mentioned another aspect about the annotation, “I also like the fact that I can look at the background, on how that clip was come about, I think that’s good. Because once I know how the program is thinking, I can better adjust my questions. Because I know if you think this way, than I’m going to phrase my question with keywords, because it looks like that’s what you’re looking for, and then those keywords will line up what I wanna to extract.” She thinks that the annotation interface helps her to phrase the questions in a more efficient way, which is an important finding of the system too.

iv. Paths as Materials

An interesting point that Subject 1 raised was the concept as reusing story paths as materials: “You know? So there’re a lot of uh, a lot of ways you can look at this, a lot of ways where, it can be a great resource, as a raw material, and also something that as been cooked and re-cooked many times by many users. I like that. “ Amazingly, the terms “cooked” and “re-cooked” reveal the fact that, through the combination and organization of paths that have been created by other users in the story world in addition to the clips or raw material, there remain many more possibilities of interesting ideas for story flow.

v. Theme for the clips and also for the paths

Based on the previous point where paths can also become the *materials* in the corpus, the users – and potentially the system itself as well – will be able to understand the corpus in a higher-level, even more storied way, than navigating through a story world which is composed of individual sequences only. Subject 1 said “...the themes could have not only all of the clips, but also the mini segments already made, I think that would be very useful. I could see that being very useful for, trying to understand what is in common with all of these difference workshops in different countries and different cultures. In so many ways they’re the same but also different.” Extending her idea, the themes of the paths may not be as those for the individual sequences (e.g. “Victoria’s learning difficulty”) any more. Instead, they will become higher level, and more similar to the “lessons” or “observation” that Schank talked about, such as “You should be yourself, not blindly following others or trying to act cool”. Indeed, a story world that has this kind of ideas flowing around because of the existence of story paths created by other people, will be really useful for the system to perform more intricate understanding processes.

vi. Through the Different Corpuses; Through the Different Times

Subject 1 talked about the videos that she worked on, which is about Prof. Seymour Papert’s lectures that he has given in different places in the world throughout his life. She pointed out that Papert talked about the same concept with different accompanied examples in different times or locations because of the different contexts. It would be really useful, she said, if the system could

pull out related video segments from different corpuses: "...I would like the system to navigate through all of the stories of each of the talks, not just one of them, and bring out to me maybe one clip that is from 1983, and another one from 92 and another that's 2007, for example. And then it's up to me to do my own collage. But I would like to know if the navigation it's possible, not within one movie that has been segmented, but between many other movies as well. ("Yeah," I responded.) Yeah it's possible? Oh that's....that's really cool (She smiled)... So that really allows you to navigate through his mind, basically, the way he's been thinking about one topic, from, let's say for the last 20 years or something like that." I was surprised with her words "that really allows you to navigate through his mind, basically", but indeed it points out the heart of Storied Navigation, and the ultimate purpose that it aims to achieve.

vii. Flexibility and Encouragement of Collaboration

Subject 1 also agreed with the system's mechanism that allows multiple users to contribute the annotation together. "It (Roballet)'s a long work and it's good to have several people doing it because several people participated in it and each had a vision of what happened. So I like the fact that you can just log in the sequences, say okay everybody contributes, you know, within two days we will have annotated all the sequences because we will need the information or resources for another workshop we're doing next month or something like that. So it's good because it *invites* collaboration."

Certainly, she did not talk about the system's advantages only. The features that she suggested are:

- i. She suggested adding a "menu" or "sampling" of all the videos in the database that to certain extent welcomes the users when they come to an unfamiliar corpus. "Like an invitation", she said.
- ii. Duration of the each video clips is an important piece of information for editors. There should be somewhere in the interface that informs the users how long each video sequence is.
- iii. Path Saving. Some of the characteristics she talked about rely on a Save function for the created paths, but currently the system does not support this function because it has been developed only to enable users to perform the tasks. Subject 1 pointed out that it is not just a saving function, but naming of the paths will bring fruitful information to the users as well, "the way you name the path, might give ideas to others on how to use this movie as a resource too".

5.1.2.2 Study 1: Subject 2

Subject 2 was one of the graduate students who participated in the sensor development in

Table 5-1: Comparison between the Two Stories Made by Subject 2

	Story 1	Story 2
Audience	Imagining showing the story to people who don't know anything about Roballet	Not Mentioned.
How to Start	Starting by a clip of the final performance, in order to show "what it is" first	Typing the sentence, "Meetings with Seymour about the general ideas of using technology in a dance performance"
Main idea	"It's not that kids only do what they are told to do, but they actively create animation, dance...That's where the learning take place"	"The second is more of a story. And there was a behind-the-scene story that I wanted to tell" "Idea, problem, problem, problem, problem, and the final performance"
Ending	"And the last one of course, just Seymour talking about the basic idea. Basically he just described the whole idea of the workshop. That's sort of an ending"	"The performance is in itself the resolution, probably."
Particular Characters	"I didn't choose any particular person. I show Louis because it (the clip) is the best one that fits the thing that I wanted to say." "A choreographer coming in...he was trying to each particular thing. Not just kids arbitrary doing things in random, but he put strong emphasize on what kind of dance as well"	"The choreographer was the conflict, the most ...I have him in mind at first..." "But at the end I used Seymour as the technology people. His role sort of emerged...Toward the end one his role was established. Then I searched Seymour specifically just to maintain the consistency of the characters"
"Storyedness"	"It's not really a story, just like an informative video about learning, Seymour talking...just many aspects of the workshop. It's not really about any particular person, or events. It gives a general idea, just a bunch of clips that I think as best representing the different aspect of it"	"This one has a little bit more of a story, right? Seymour was excited, the dancers are excited, everyone was excited, but then the technology doesn't work! Everyone was frustrated: Why is the thing so slow? Things are not working the way they should...?" "You know...I was motivated. And I was fully aware of the difference"
Functions Used	Mostly Keywords: Seymour Papert, Lights, sensors, etc	"This time I use the mostly the characters search and the story structural role, because I want to look for rise and conflict...and that was helpful." "I sometimes just put in a word "programming" not because it's too slow, like a Google search, because I have the best idea of the best clip but it doesn't pull out the one that resonated with it" "See other searched results is very nice, because I wanted to see other ones that were pulled out by the text." "I didn't wanted to see the alternative, I didn't use emotion...um...I did use it once, but I found that using story structural role was a lot useful. I guess I use conflict. I left that on, and I search with rise and resolution. Because I wanted to see when it started, right? The rise seems to be more accurate...because I guess resolution doesn't really exist for this type of story."

Roballet, and also has 50-100 hours of experience of video editing. His most familiar editing system is Adobe Premiere. To build the first story, he spent about 36 minutes, and he spent 32 minutes for the second one. Since, unlike the previous subject, he made both of the stories, I will emphasize differences between the two stories as well as similarities.

Table 5-1 shows the comparison between the two storytelling activities that Subject 2 participated. The difference between the descriptions that he read before making each story is that, the second description gives a hint of controversy. That is, two different perspectives did exist in the workshop, similarly to what was described in this passage: the “technology-mania” and the “art-mania”. Correspondingly, the first story that Subject 2 made was “not really a story, just like an informative video about learning”, as shown in the “Storiedness” column, as opposed to “This one has a little bit more of a story, right?” which he said as the very first sentence after the second story was made. He said “I was motivated. And I was fully aware of the difference”, emphasizing how the different directions influence the storytelling tasks.

Also, from the functions used in the second story, from the two pieces of explanation in Story 2, “I have the best idea of the best clip but it doesn’t pull out the one that resonated with it” and “The rise seem to be more accurate”, we can find that he was much clearer about which specific sequences he wanted to, as opposed to the uncertainty shown from the description of the particular characters in Story 1, “I didn’t choose any particular person. I show Louis because it (the clip) is the best one that fits the thing that I wanted to say.” From here we can derive a hypothesis:

Hypothesis 1:

“When a storyteller is more motivated and clearer about the story that he/she is to tell, this storyteller will be more specific about the target material that he/she wants to use to construct the story.”

In the “Function Used row”, Subject 2 also described the system as a “search tool” much less hesitantly after making Story 2 – not only referring to the edit-by-typing function but also search by character as well as story structural role, which are not necessarily similar to what “search” typically means. If we combine this finding with the above hypothesis, we may derive a second hypothesis:

Hypothesis 2:

“When a storyteller is more motivated and clearer about the story that he/she is to tell, the functions that he/she needs in the storytelling process will be more direct and less suggestive”

In other words, under the circumstances where the storytellers are more motivated and clearer about the stories that they have in mind, they will use the search functions in targeting their story materials more than using recommendations such as “What’s Next” or “Find Similar Alternatives”, because the results responded by the later are relatively arbitrary.

Furthermore, in the “Particular Characters” row, Subject 2 said, “...at the end I used Seymour as the *technology people*. His *role* sort of *emerged*...Toward the end one his role was *established*.” We can find that the second story has more sense of “position” of the characters. That is, these characters have more perspectives and irreplaceable statuses, because they carry particular meanings or functions in the story. Based on this finding, we can make a series of inferences as below:

- i. In Subject 2’s second story, the characters are established along the way
- ii. Characters are established along the way even when the story idea is clear
- iii. The storyteller may be looking for the establishment of the characters or perspectives in the storytelling process when they have a clear story idea
- iv. The recommendation the storytellers need may not be finding what’s next or alternatives of the existing sequences, because they are looking for the characters that represent the perspectives which no one represents yet in the story
- v. What they need may be a tool for searching characters based on perspectives
- vi. In order to provide recommendation based on perspectives, each character needs to be identified with his/her perspective
- vii. If the system can provide a tool for the storytellers to search for characters based on the perspectives, they will find establish the characters for their story ideas that are clear in their mind more easily.

This gives us a suggestion of modifying Hypothesis 2. Hypothesis 3 shows the revised version of Hypothesis 2:

Hypothesis 3:

“When a storyteller is more motivated and clearer about the story that he/she is to tell, the functions that he/she needs in the storytelling process will be more direct, yet more complex than the currently provided search functions in Storied Navigation. For example, search for characters by perspectives.”

Another hypothesis may also be derived from the increased usages of “search by story structural role” in telling the second story:

Hypothesis 4:

“When a storyteller is more motivated and clearer about the story that he/she is to tell, he/she will make more consideration for the overall story structure, while considerations for the sequence adjacency may remain the same”

This hypothesis can also be supported by the “Main Story” row, which shows that Subject 2 had more sense of the overall structure in the second story, while in the first story he only had the idea of what he will show in the stories, not necessarily their order. That is, he tried to initiate Story 2 with an exciting idea, followed by a series of problems, and end it with a performance, which somehow serves as a resolution to the problems. “The rise seems to be more accurate”, he mentioned his experience about searching by story structural role, with the combined criteria of “conflict” with either “rise” or “resolution”, “because I guess resolution doesn’t really exist for this type of story...The performance is in itself the resolution, probably.” To me, what he meant was, the characteristic of resolution is not depicted on the surface of the last performance sequence, because it does not suggest any existence of conflict. What it *does* suggest, however, is a sense of convergence, which could be used as a type of resolution, depending on the story it is used for. Therefore, we can conclude that a “more *storied* story”, like Story 2, requires more complex “ways to think” for arranging the sequences, and is more difficult for the system to provide good recommendation. Hypothesis 5 gives a more concise summary of this finding:

Hypothesis 5:

“When a storyteller is more motivated and clearer about the story that he/she is to tell, the story structural role of a sequence may depend more on the story it is used for.”

To conclude the five hypotheses that I listed above, I propose two functionalities for the next version of Storied Navigation:

For storytellers who are more motivated and clearer about the stories they are to tell, they will need a system that 1) makes accurate inference about the story structural roles that a sequence possibly has, by trying to understand the stories that they are trying to build along the storytelling process 2) provides more storied search functions that allow users to search by more intricate criteria, such as perspectives.

Overall, Subject 2 thinks Storied Navigation is a search tool. He said, “It allows me to search the database. It matches my need.” Whether using the system “as a Google search start with long sentences,” for the first story, or “also search the conflict and resolution” for the second, to him it is a search system for storytelling. In addition, he mentioned that it “should be a nice learning tool too...especially for kids, right?” which provides a possible direction for further investigation as well.

5.1.3 Study 1: Summary

In this subsection I reported and how the two subjects in Study 1 reacted in completing their tasks, and did some analysis of my observations. Generally, the two storytellers succeeded in finishing their stories and enjoyed their storytelling activities, according to both the questionnaire and the qualitative analysis. Seven characteristics of Storied Navigation were pointed out by the first subject, namely:

- i. Improving Efficiency for Documentary Video Editing
- ii. Quick Annotation. New Way of Organizing the Data
- iii. Annotation Interface Helps Understanding of the Logic Behind
- iv. Paths as Materials
- v. Theme for the Clips and also for the Paths
- vi. Through the Different Corpuses; Through the Different Times
- vii. Flexibility and Encouragement of Collaboration

And, five hypotheses were made according to the comparison between the two storytelling activities participated by Subject 2. Based on these hypotheses, I conclude with two directions for further development of the next-generation Storied Navigation system:

For storytellers who are more motivated and clearer about the stories they are to tell, they will need a system that 1) makes accurate inference about the story structural roles that a sequence possibly has, by trying to understand the stories that they are trying to build along the storytelling process, and 2) provides more storied search functions that allow users to search by more intricate criteria, such as perspectives.

5.2 Study 2: Using the System to Browse Videos

5.2.1 Study 2: Design & Setting

The goal of this study is to investigate whether the design of the Storied Navigation system serves as a solution to the other problem proposed in Chapter 1: “How can a system help users to gain familiarity with the materials stored in the corpus?”. More precisely, I want to verify two hypotheses: 1) “*the system is helpful in the process of browsing an unknown video corpus*” and 2) “*when the subjects gain better understanding about how to use this system, they will browse the corpus in a more storied way*”. Both qualitative and quantitative analyses are provided for this goal, because there are too many attributes that contribute to the “storied” characteristic, which makes a purely quantitative analysis difficult in a short period of time.

The design of the study is as follows. The subjects are all unfamiliar with the Roballet corpus, and should all have experience of browsing videos online. There are two sessions, in each of which the subjects are asked to browse videos in the Roballet corpus using the Storied Navigation system, similarly to using YouTube, and I try to observe the difference between the ways they use the system. After each session, the subjects are asked to talk about what they see in this session, what kind of character emotions they found, and what kind of story they would tell based on the experience they have in this session if they were to tell a story about Roballet. The purpose of the discussion is to encourage reflections in their mind of how the storytelling process proceeded, why they made the decisions they made, etc.

Strictly speaking, there are three criteria that may affect the result, namely, 1) different degrees of the subject's understanding the Roballet corpus, 2) different degrees of the subject's understanding the Storied Navigation system, and 3) the discussion and reflection after the first session that might influence the second session. I assume that the different degrees of understanding the whole corpus will not influence the results greatly because they have different focuses. For the discussion and reflection between the two sessions, I assume that can be included as part of the understanding of the system, because in the first session it tends to be the case that they don't know how to really take advantage of the functions, even know they do know how to use those functions. As the reader can see in the later subsections, although I was not able to conduct a more ideal study that controls more criteria, there are still interesting and valuable results from the observation of this study.

The procedure of the study goes as follows. After, providing their general information, the subjects will go through an introduction of the system's functionality, similarly to the previous study. Each of the instructions contains a few examples that they are asked to follow, while the content of the examples is as irrelevant to the later task as possible. After the instruction, the subjects are asked to do one warm-up example, and proceed to two browsing tasks. In every browsing task, the subject will be given a key story, similarly to the previous study, while the key story is used only to provide the initial context in their mind, not a goal story that they need to try composing.

Example:

Henri is one of the children in Roballet. He became directive toward other children in the workshop, and did a lot of other things, including making animation, dancing, and so on.

The two key stories are the learning story (Story A) and the "technology vs. art" story (Story B). To eliminate the effect to the result caused by the order of the two stories, half of the subjects (three subjects) used the "learning" story as the first key story in our study, and the other subjects (four

subjects) started with the "technology vs art" story. Each browsing session has a 25-minute time limit. The subjects can browse the videos in their own pace, using whatever functions they like. Below are the two key stories:

Story A:

Roballet is a workshop where the researchers try to investigate new ways of learning. That is, learning can take place by creating things, by emulating others, by collaborating, by problem solving, by teaching, by planning, and so on. Please use "Learning in Roballet" as the main topic, find video sequences that you find interesting. You can start by focusing on particular aspect of learning if you like, or you can start by focusing on certain particular characters too.

Story B:

Roballet is also a workshop where art is created by introducing technology into the creating process, such as using computers to create animations, using sensors in dance, and so on. In addition to these practical activities, there are also many discussions surrounding the topic "How technology and art can or should be combined in the workshop". Please use "Technology and Art in Roballet" as your main topic to find video sequences that you find interesting. You can start by focusing on particular activities, or you can start by focusing on certain people too.

Note that Story B is different from the other Story B in Study 1, where the sentences suggestive of the existence of conflict or different perspectives are removed, such that the two key stories will provide information of similar quality and quantity.

5.2.2 Study 2: Observation & Results

There are seven subjects in the first study, the storytelling tasks. All of them are "not quite familiar" or "not familiar at all" with Roballet and have experience of browsing videos online. According to the questionnaire on the 5-point Likert scale filled by the subjects in this study, all of the subjects "agree" or "strongly agree" that they can use the system to browse videos, that the system helps them to find what they want, that they can use the system to edit videos – even though they were not ask to edit videos in the task and three of them had no experience of using any video editing systems, and that the interface design is "intuitive" or "very intuitive". Five out of the seven subjects think using the system is "enjoyable" or "very enjoyable", and all the subjects who have video editing experience agreed that using the Storied Navigation system is "easier" or "much easier", comparing to using other video editing systems.

Table 5-2 shows the usage frequency of the functions available to the users. The “Search by Keyword” is a function that searches exact-matched keywords in the story descriptions in the videos’ annotations, provided together with other search functions in the interface. The “Reposition” function, on the other hand, stands for the drag-and-drop action that the users perform in order to change the sequential order of the sequences in the timeline. All the other functions have already been introduced in Chapter 3 and Chapter 4. The second column indicates the number of subjects who have ever used the respective functions in the whole study, and the third column stands for the accumulated usage of the respective functions by all the subjects. The green boxes highlight the functions that are used by many subjects and for many times, whereas the red boxes highlight the functions which no subjects ever used at all. Below I discuss the subjects’ opinions related to these highlighted functions, including Edit-by-Typing, reposition, and search.

i. Edit-byTyping

Among all the functions listed in Table 5-2, Edit-by-Typing is the most frequently used one. Nevertheless, different people used it in different ways, under different circumstances. One subject said, “I started with the edit-by-typing function, but then it seemed that it’s hard for me to make long sentences...and it takes so long for the result to come up, so then I just try to look for things that I wanted from the result it gave me.” On the other hand, another subject said, “I searched the related keywords: sensor, animation, dance...but then I thought it’s too slow to do it one at a time, so I decided to type a whole sentence to make a sequence...so I started to type here”, and one subject also said, “the problem of the sentence search is that it matched the words that I didn’t mean it to use to match...and the result are often different to what I expected...it especially likes to mark the word ‘good’, but I didn’t mean that in my search” These tell us the pros and cons of the edit-by-typing function: it gives you flexibility of describing relatively complex meanings that keyword search cannot accommodate, but it also takes more time to process, and the matched terms may not be the focus. It also explains why some people became more willing to use this function, while others became more reluctant, as shown in Table 5-3.

ii. Reposition

Repositioning of the video sequences in the timeline is somewhat interesting. The subjects were asked to find unfamiliar but interesting sequences and were encouraged to put these sequences in the timeline, but I did not ask them to make a nice order of those sequences at all. However it became one of the most popular functions in the task, as Table 5-2 shows. This suggests the possibility that the subjects have a sense of story flow even during the process of simply “finding” videos that interest them. One subject said in the interview, “Oh, I actually wanted to make a story,

Table 5-2: Functions Used in Study 2

Function Type	# Subjects Ever Used it	# Total Usages
Edit by Typing	6	52
Search by Theme	5	7
Search by Keyword	7	34
Search by Character	2	2
Search by Emotion	3	8
Search by Location	1	1
Search by Date	0	0
Search by Story Structural Role	3	10
Find Similar Alternative	4	7
Find What's Next	3	13
Reposition	5	44

Table 5-3: Changes of the Subjects with Respect to the Functions in Study 2

Function Type	# Subjects Ever Used it	# Subjects Increase Using it	# Subjects Decreased Using it
Edit by Typing	6	3	2
Search by Theme	5	4	1
Search by Keyword	7	4	3
Search by Character	2	1	1
Search by Emotion	3	2	1
Search by Location	1	0	1
Search by Date	0	0	0
Search by Story Structural Role	3	1	2
Find Similar Alternative	4	0	4
Find What's Next	3	1	0
Reposition	5	2	3

you know? Both of them...but it isn't necessarily a story with a good beginning or ending or anything...just a collage of the related concepts...putting together the things that are related to learning, and gives a final presentation of the result...at first there's a exploration, and they're not sure what to dance...I already made it a story. Yeah, I remember that...But why? It's kinda weird... Maybe it's just a habit I guess, when you write something you take care of the continuity..." The word "write" that he used was very peculiar. Our subjects have already regarded the activity as a "writing" task, which I think will not happen in using other video editing or browsing tools. It suggests that the system is really directing them to think in a storied way. Another subject said he re-arrange the order of the sequences because he wanted to "place it as the main topic, which is to find sequences that you find interesting", because "it makes me feel that maybe I can do some little adjustments to the search result, so that I can make the flow of the video match better to the technology and art topic...I think it would give a better presentation of how I think such a technology and art topic should be presented if I change the order a bit.", which is another evidence of the *storied* thinking that the subjects performed in this browsing study.

Table 5-4 Experience Descriptions about the Search Functions

Subject	Experience Description	My Observation
Subject A	"...mostly the search story structure thing. Because the other ones, like date or location I don't really care...if I keep using the same theme or characters it'll be kinda boring...so anyway I just searched with the idea of making a story...and I just wanted to find an ending or something different"	Preferred criteria: story structural role Not preferred: date, location, theme, character
Subject B	"I like the curtain, because it's pretty interesting, right?...but then I wanted to see how they came up with the dance, so I did search 'the curtain', but it didn't show anything the I expected...do you have the function 'what's previous?' or something like that?"	Used criteria: keyword Expected sequence unfound using: "the curtain" Function wish to have: "What's Previous?"
Subject C	"At first I searched 'sensor' (keyword), cuz I thought it's related to technology and I wanted to see what they used the sensors for...but then I almost watched all of them one after another...most of them were about the sensors, but I still couldn't figure out why they used it...so then I search 'art' (keyword) but I didn't find anything"	Used criteria: keyword Expected sequence unfound by using: "sensor", "technology", "art"
Subject D	"...I guessed I used 'character' and the 'story structure'...I looked for Peter because I wanted to see who he was, but it seems like he's not an important character because there's not much of his stuff...mostly beginning, ending, and converge (for 'story structural role')"	Used criteria: character, story structural role Expected sequence unfound by using: "Peter"

iii. Search

Different subjects used the search function with different criteria. Table 5-4 gives a clearer comparison between the subjects' experience with the search functions. Briefly, the users who used the story structural role had good results, while keyword search did not necessarily give them useful results. According to my observation, the problem of keyword search may be due to the fact that they used keyword search when they were not sure how to specify their needs, or, equivalently, there were no ideal criteria for them to use. Thus, they turned to what they feel the most familiar – keyword search. If the system had functions that were closer to their needs, for example, "What's Previous?", or implement the idea of finding related topic taking place in a previous time of the story into existing functions like "What's Next?" or a question-answering scheme, there is possibility that this problem might have been resolved.

In addition, according to Table 5-2 and Table 5-3 we can find that the most popular search criteria were keyword and theme, whereas the date and location were the least preferred. According to the interview, the subjects were reluctant to use date and location mainly because they were unfamiliar with the event, so the time and location were not as meaningful to them. One subject, though, did mention that he kept using location as the criteria for the "What's Next" function, because he wanted to focus on how the children made their animation. Therefore, whether date and location are really unhelpful or how the functions related to these two criteria should be modified, will need further investigation.

Overall, the subjects regarded the Storied Navigation system as a search tool for people to find what they want, while it is largely different from Google search. Here are their reasons:

- i. The search functions help people to specify the target that they want by narrowing down their rough ideas
- ii. The annotation (including the story description, the theme, the location, the parsed representation, etc) also clarifies "other people's perspective, how they look at it" as well as how the system reason these sequences, and, accordingly, how users should modify their way of search
- iii. The system is helpful in dealing with large amount of data, because it allows users to organize the found data directly in the search interface, such that they won't need to jump back and forth between the search interface and their working environment, which one subject found very important
- iv. The experience is fun because it is a collective activity, as one subject said "it's like online shopping. You can put what you want in the shopping cart"
- v. It can be a new way of blogging – a combined version of textual and video blogging, because the experience of searching and arranging sequences is "not like introducing a

- restaurant to a friend. But it's like you digest it a little bit, and then present it to them"
- vi. The "What's Next?" function is very helpful, as one subject said "it's really useful. YouTube really should add this in their website."

On the other hand, their suggestions or criticisms of the system include:

- i. It is not clear how to use the criterion, "theme".
- ii. The video corpus is not big enough
- iii. Edit-by-typing is slow.
- iv. Edit-by-typing may match the words that the users find less important in the sentence
- v. A few subjects found the followings not as useful: "find similar alternatives", "search by date", and "search by location"
- vi. Some subjects did not like the video corpus, because they do not like children or other reasons

Finally, an interesting thing to mention is that, some subjects focused on the textual annotation of a video sequence to determine whether they wanted to spend more time with it, while others focused on the thumbnails. This varies from user to user, but which ever they choose, according to my observation, it tends to be extreme – many of them either totally ignore the annotation or vice versa in the searching process. Whether the interface should be designed with regard to the annotation versus the textual capture will require further investigation as well.

5.2.3 Study 2: Summary

In this subsection I report how the seven subjects in Study 2 reacted in completing their tasks, and did some analysis for my observation. Generally speaking, this study shows that the users agreed that the system is helpful in finding what they wanted in their browsing activity, according to both the questionnaire and the qualitative analysis. The functions that they used the most include "search by theme", "edit-by-typing", and many of them found the "What's Next?" and "search by story structural role" useful. These are all novel features in the Storied Navigation system.

I did not succeed in comparing the two browsing tasks within individual subjects, because the data that I collected did not indicate any comparative information. There were also complaints about various things in the system design, including the speed or efficacy of certain functions, but according to the presented results, the usability as well as the overall efficacy of the system are established.

5.3 Statistics of the Questionnaires

There are two sets of questionnaires, each for one of the studies, and there are four different questions between them. The question that is only in the questionnaire of Study 1 is: “In the experiment, I was able to use the system to edit videos”, where both of the subjects chose “strongly agree”. The questions that are only in Study 2’s questionnaire, are “In the experiment, I was able to use the system to browse videos”, “I can use this system to edit videos”, and “ The system helps me to find what I want”, where all the subjects chose “agree” or “strongly agree” as their answers for all of them. Figure 5-2 shows the 19 questions that co-exist in the two questionnaires, where the parentheses indicate the options types in the 5-point Likert scale. For example, (“easier”) indicates that the options that the subjects can select are “much harder”, “harder”, “neutral”, “easier”, and “much easier”, whereas (“agree”) stands for “strongly disagree”, “disagree”, “neutral”, “agree”, and “strongly agree”. Table 5-5 shows the result of the questionnaires. The green bar stand for the results of 5 points (“strongly agree”, “very enjoyable”, “very useful”, “much easier”, “much better”, or “very intuitive”), the yellow bars stand for 4 (“agree”, “enjoyable”, “useful”, “easier”, “better”, or “intuitive”), the orange and red bars for 3 and 2, respectively, and the white bar in question 14 stands for “Not applicable”, because they have no experience of video editing.

Result of the Questionnaire

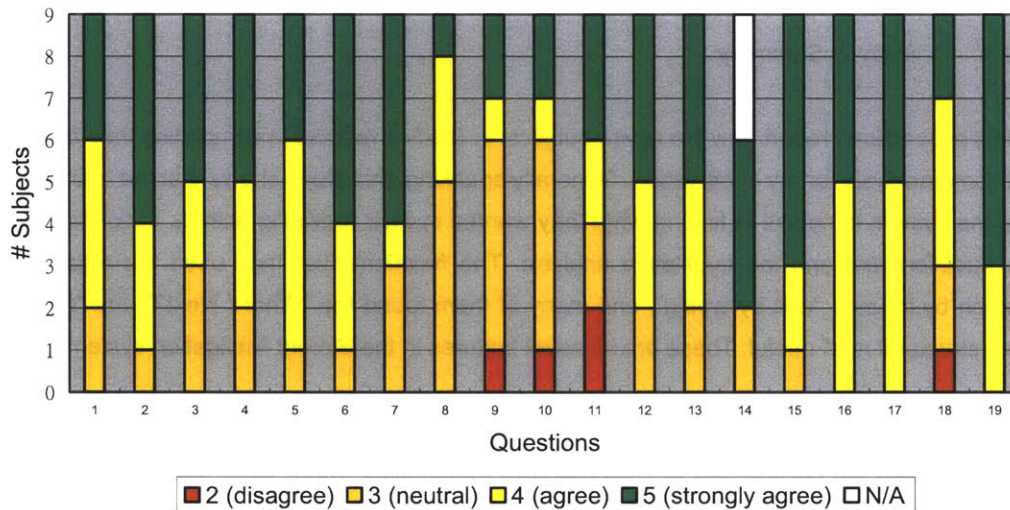


Figure 5-2 Number of Subjects who Rated 4 or 5 Points to the Questions

From Figure 5-2, we can see that most of the questions are agreed to by over half of the subjects, except for question number 8, 9, and 10. For question 8, three subjects found it “very useful” while others do not, which gives us the information what a few subjects used and liked this function

particularly more extensively than the others, considering the result in Table 5-2 and Table 5-3. Then, according to question 9 and 10, not many subjects found “search by location” and “search by date” useful, which matches the result in Table 5-2 and Table 5-3. What is interesting is, there were some subjects who did think these two functions are useful even though they may not have used it. According to their explanation, it is because they think it would be useful if they are using the Storied Navigation to browse a larger video corpus, “like watching the drama at YouTube.”, for example.

Table 5-5 Questions in Questionnaires Study 1 & 2

1	In the experiment, using the system is (enjoyable)
2	I find the “Edit-by-Typing” function (useful)
3	I find the “What’s Next” function (useful)
4	I find the “Find Similar Alternatives” function (useful)
5	I find the “Search by Keyword” function (useful)
6	I find the “Search by Character” function (useful)
7	I find the “Search by Emotion” function (useful)
8	I find the “Search by Theme” function (useful)
9	I find the “Search by Date” function (useful)
10	I find the “Search by Location” function (useful)
11	I find the “Search by Story Structural Role” function (useful)
12	I find the video detail (annotation) interface (useful)
13	I find the yellow box that shows the matched text and annotated stories of videos (useful)
14	Compared to using other video editing systems, using this system is (easier)
15	I would like to have this system in my computer to organize my personal videos (agree)
16	I would like to have this system as a online collection of my personal videos as using Flickr or YouTube (agree)
17	I would like to use this system to browse videos on line (agree)
18	Comparing the browsing interface of online video websites (for example, YouTube), the interface of this system is (better)
19	Overall, the interface design is (intuitive)

5.4 Summary

In summary, this chapter reports the result of the two conducted studies that were designed to verify whether Storied Navigation has provided solutions to the main problems introduced in Chapter 1: 1) “How can a system help users to develop their story threads?” and 2) “How can a

system help users to gain familiarity with the materials stored in the corpus?”. From the qualitative and quantitative analysis of the subjects’ activities as well as the filled questionnaires, I conclude that the system has successfully provided two solutions to these problems. I also derived seven characteristics of Storied Navigation, five hypotheses and one concluded future direction from Study 1, as well as the comparison between individual functions in Study 2. Although there were suggestions or even complaints about certain aspects of the system, and it is unfortunate that I could not find strong evidence indicating the comparison between the first and second tasks in Study 2, from the two studies we can find that the system is usable and effective, in assisting the users in both storytelling and browsing activities of a video corpus.

VI Discussions

“Storytelling, we may conclude, then, is never neutral. Every narrative bears some evaluative charge regarding the events narrated and the actors featured in the narration” – R. Kearney, 2002.

In this chapter, I would like to talk about this thesis project more from a filmmaker’s or storyteller’s perspective. I will make several discussions by surrounding this central topic, Storied Navigation. Most of them compare it with other storytelling forms, e.g. filmmaking and other existing work of interactive storytelling, with regard to different aspects including *syuzhet* and *fabula*, the story path, the guidance or recommendation given by the systems, the storytellers’ level of responsibility, and so on. I hope, from a different view of looking at this project, the readers can gain a broader and deeper understanding of how I have been motivated, and how the project or even the whole field is expected to be advanced in the future.

6.1 Syuzhet, Fabula, and the Story Path

Meaning emerges as juxtaposition takes place. My mentor Glorianna Davenport wrote not too long ago, “Video stories are constructed by selecting and ordering one or more shots and sounds into a meaningful sequence. Sometimes the editor is seeking the next shot in a sequence. Sometimes she is selecting the shot that will set up and begins a new scene. The editor is always concerned with what arrangement will create the right meaning.” [47]. It is the editing process that *makes* the stories, because different meanings will emerge as the editor put together the sequences in different temporal orders, and these meanings give the audience different perspectives toward an event, a person, or the whole spirit, whole lesson of the overall story. The set of meanings communicated in a film are changed over and over again as one after another cut is made, and both the *syuzhet* and *fabula* are re-created as well. The word *syuzhet* means the temporal order that the event and characters in a story are narrated, whereas *fabula* stands for the inter-relationships between characters, between events, and between characters and events. Certainly part of the *fabula* comes to exist (in the form of raw footages) as soon as the shooting process begins, yet a major part of it, the part that depends on the *meanings* created via juxtaposition, will not emerge before the editing activity takes place.

In the process of editing a documentary film, the most commonly asked question to the filmmakers after the test screenings is: “What is the story?” In his book *Writing, Directing, and Producing Documentary Films and Videos*, Alan Rosenthal also wrote, “*Story*. This is your first and most

impelling commandment. Find the most compelling story in your material. A philosophic essay will gain you brownie points, but a compelling story will gain you viewers. What is your story and where is it going?" [86]. Truly, even the filmmakers themselves are looking for the stories. They look for the stories by starting with a script, a pre-mature story with some uncertainty, and then making one after another cut, or one after another syuzhet. By watching the cut and examining the syuzhet they have made, they start to question themselves whether the fabula behind the scene is what they wanted to present, and, accordingly, come up with a new, modified version of the fabula, which results in the next syuzhet, or the next cut. Therefore, to documentary filmmakers it does not make much sense talking about fabula and syuzhet to them, because they cannot separate them in making a film.

An interesting thing is that, this inseparable characteristic of syuzhet and fabula is somehow applicable as well in the scenario of Storied Navigation. This is because, what the audiences do under this scenario is actually very similar to the editors' tasks in filmmaking – in particular the creation of the stories. That is, the users are making the fabula and syuzhet for their stories at the same time too. Certainly, for a self-contained story project, like the Roballet corpus used in Chapter 4, a big part of the fabula always remains the same because the number of usable clips is limited. But consider a scenario where the user can use all the video clips in YouTube. It is very likely that this "big part of fabula" will surprisingly vanish because the amount of the usable material grows drastically, and there is no way one can expect what a video story built by an arbitrary user will come out to be. In this case, in my perspective, the fabula and syuzhet will become one integrated form – the navigation path. It will be the *path* that becomes what the *story* is, and meanwhile what the *telling* of the story is.

The path is not only the story and the telling of the story. The editor of *The Godfather Part II* and *Part III* Walter Murch wrote, "My point is that the information in the DNA can be seen as uncut film and the mysterious sequencing code as the editor." [82]. In the Storied Navigation activity which can be regarded as an emulation as documentary editing, the story path can also be viewed as a product of the user's "DNA". That is, it will be the path that incorporates the storyteller's perspective because the path is formed based on the storyteller's interests or likes/dislikes; it will be the path that presents the particular meanings of the individual juxtapositions to the storytellers because sequences are juxtaposed based on the storyteller's instinct of what should come before or after what; and it will be the path that reflects what the storyteller was touched by, what the storyteller has learned from, and what how the storyteller has communicated with, because of all the arrangement that can possibly be resulted, the storyteller chose to make the path in the way he/she did.

6.2 The “Accompanied Editor”

To make a comparison between Storied Navigation with another thing – conventional work on interactive storytelling, I think we can say that the users navigate the story world as if there was an “accompanied editor” that does not exist in any of those conventional interactive storytelling schemes. That is, in the previous works, either no path guidance or recommendation is provided [13], the story domains are confined (i.e. the systems do not support navigation in corpuses general story materials) [48, 49], the mechanisms rely on keyword match and do not perform any other sorts of understanding of the stories [29, 30], or the paths are pre-determined, which means that the users can only develop their stories by following a small set of designed paths instead of receiving recommendations based on arbitrary existing paths that they have made [76]. Storied Navigation is the first work that, for general story materials, acts as an “accompanied editor” that looks at what the users have in their timelines and proposes suggestions that might serve as good ways of growing the stories, just like a human editor that sits with the users side by side. That is, whenever a user wants to find something next, the way he/she interacts with the system is similar to interacting with an experienced editor – asking “what do you think we can add next?” and getting several suggestions to adopt from. Sure what a real human editor can suggest greatly exceeds what the current Storied Navigation system can do. Nevertheless, I think one of the lessons that we may learn from this thesis project is that, we should really start thinking about the “ways to think” that film editors have and perform, and start implementing them into systems such that they will be able to guide the users through the navigation of media collection, so that the media elements will be “fabricated” via stories for users to view, share, and regain their memories from, not just as fragmented pieces as what they are today.

6.3 Goal

In order to do so, we need to talk about the topic of *goal*. A lot of technical papers talk about using goal structure for automatic narrative generation, based on a theory that the story is told in order to achieve certain “story goal”, and storytelling or generation is a task of goal recognition and plan generation. I think it is reasonable to formulate the problem this way, because as a filmmaker I did have goals that I wanted to achieve in making *Life. Research*. My goal was to make a portrait of Hugh, to reveal the fact that the accident did change his life and his belief and attitude towards life, and to bring to the audiences’ faces the connection he and I built with each other throughout the filmmaking process. To me, stories do emerge with a set of pre-determined goals in the tellers’ mind.

The storytellers’ goals in the Storied Navigation activities, however, are somewhat different to those in traditional filmmakers’ minds. The number of formulated goals can be plotted as a point on a

line, where the left end of this line is 1, and the other end stands for the queued sequences, or the length of the path. The location of this point varies with different user, different story theme or domain, etc. At one end of this line, similarly to the hypothesis in those previous works, the whole story has a cohesive goal that it aims to achieve, for which each of the material selections and juxtapositions is made. At another extreme, the goal may be changed or initiated constantly throughout the navigation process, such that each sequence selection in itself has its self-contained goal that may not constitute any collaborative goal with the any of the other goals. In this later situation, which may happen for most of the time, I don't think any goal structure on the level of the "happening of events" that captures causality or even chronology would work for Storied Navigation, because the focus of the navigator flips rapidly from one character to another, from one location to another, and so on. Nevertheless, I do agree that this kind of goal structure approach could be useful if the goals are higher level ones that capture the lessons the viewers may learn, or the thoughts they may reflect upon as they navigate the story corpus. How to formulate a computational representation for these high level goals, though, remains a problem that I haven't seen any of these work addressing.

In his book *In the Blink of an Eye*, Walter Murch wrote, "*How do you want the audience to feel? If they are feeling what you want them to feel all the way through the film, you've done about as much as you can ever do. What they finally remember is not the editing, not the camerawork, not the performances, not even the story – it's how they felt.*" [82]. "How they felt" is so important to a film that the possibly best editor alive in the world would make such a strong statement – even more important than the story! It is not simply emotions of individual characters in individual scenes *feel*, however. What Murch meant by "how they feel", in my perspective, is a whole body of mental reactions taking place in the audiences' mind comprised of the lessons, inspirations, thoughts and reflections they derived from the story, which I think match the high level story goals that we mentioned. In other words, it is the lesson, inspiration, thought and reflection in their minds that the audience will take away with – not the editing, camerawork, performances, or even story of the film. Therefore, if we can come up with appropriate representations that allow computers to manipulate with these goals, it is very likely that we can refine the recommendation scheme to a more cohesive way where the system will have the sense of a whole story, not just selections for individual juxtapositions based on a set of regional criteria.

6.4 Lessons, Inspirations, Reflections

Finding such representations is not a simple task, however. To investigate what functionalities these representations will be used to facilitate, we can again refer to Roger Schank's book, *Tell Me a Story*. Schank talked about how *storied thinking* is important in human intelligence. He listed the following thinking processes that a storied mind should be able to perform:

- a) memorizing (storing things in the form of stories, i.e. using story elements as indices to store information.)
- b) being reminded (retrieving things from memory in the form of stories)
- c) comparing knowledge (finding similarities and dissimilarities among stories)
- d) doing analogy (find stories that share similar indices in all kinds of realms)
- e) understanding or comprehending (proposing casual relationships between observed phenomena and certain explanations by connecting them as sequenced story)
- f) planning (building up plans for certain goals by sequencing stories)
- g) creating hypothesis or theories (doing generalization based on stories)
- h) problem solving (building up solutions for problems by analogy, hypotheses and plans using stories)
- i) giving examples (find one or more stories that possess the indices or story elements described in a theory or generalized idea)
- j) explaining theories (give analogy, comparisons, examples, and plans of theories to communicate with other people)
- k) articulating (for one story, find from many of the possible theories one theory that possesses certain indices which satisfy certain criteria, and, optionally, explain this theory)

According to Schank, these are important aspects to a good theory of human mind, and are crucial to his interpretation of the suitcase word "intelligence". And, all of these *storied thinking* ways rely heavily on a good indexing strategy for stories, which is the reason why he proposed the indexing technique that we mentioned in Chapter 2.

In the theory and system design chapters I introduced how we can extract features from plain story descriptions like character, emotion, theme, date, location, etc. To me, these relatively lower level features for useful in performing the less complex ones of the above thinking ways, including memorizing, being reminded, comparing knowledge, and maybe even performing analogy. General planning and problem solving requires big enough knowledge base that contains hierarchical information of plans and sub-plans as well as goals and sub-goals, and I believe they are also achievable with modern computer science technology if we had such a big knowledge base. To perform the more complicated storied thinking ways that Schank proposed, i.e., understanding or comprehending, creating hypothesis or theories, giving examples, explaining theories, and articulating, we will need computational representations for story indexing that are higher level, such as lesson, inspiration, and reflection that we have discussed about.

One possible way of representing these indices is by using other stories. Suppose we use series of stories to represent a lesson that other stories might implicitly suggest. In order to determine the lesson for a story description S_1 , we could perform analogy between S_1 with a considerable

amount of others that already have specified lessons, and find the best possible lesson shared by the analogous stories statistically. To me this is a reasonable strategy because analogous experiences tend to share the same lessons, and it is very likely that we could find these lessons in this thesis project if there were big enough knowledge base of lesson-specified stories that can be used to perform analogy. Once we can derive the lesson for an arbitrary story, and even other high level story features like inspiration or reflection in the same way, it would be possible for the system to recognize what type of lesson or inspiration the user may be unconsciously looking for, and to give more profound guidance in the navigation process. Under that scenario, I cannot imagine how intriguing the navigation experience will become, because of the endless amount of possible story paths resulted from the corpus' huge capacity, and the navigators' high anticipation of the recommendation – “What’s Next?”.

6.5 The “Democratic” Interactive storytelling

In the context of traditional cinema, filmmakers are the ones that are responsible for the stories the audience receive. Their mission is to find the *right* stories according to their experience, their taste, the main message that they aim to deliver, the market, etc. It is a difficult task, because at times to time these criteria may contradict with each other. For documentary filmmakers, the film may need to reveal the protagonist's innermost feelings, while the person being filmed may find the filming process intrusive. When a filmmaker wants to celebrate the great achievement of the protagonist, everyone in the interview may keep talking about bad things that this person did. How to make the right choices is always the biggest challenge to the filmmakers. After all, it is the right choices that make the stories successful.

On the other hand, in the world of Storied Navigation, the stories are created by the users. That is, to certain extent the audiences themselves have become the storytellers, and the providers of the media collection who capture the videos with their camcorders will not need to figure out the most desirable navigation path for the viewers. Consider using all the videos on YouTube for Storied Navigation, there may not be any providers that own most of the media files, because the materials are contributed by the public collectively. The audience will need to be responsible for what they see, how they feel, and what resonates in their minds, because the story emerges when the navigation takes place. An audience may not build a *good* story path from a professional's perspective, but may enjoy great pleasure from the process of exploration, which traditional cinema hardly offers. Different audiences may also come up with different stories even when surrounding the same character or same topic, because a good leader to an audience may seem bossy to others and therefore leads to two different stories, for example.

In other words, Storied Navigation is a more objective, more “democratic” medium; meanwhile, it

relieves the filmmakers from the “obligation” of finding the *right* story. It has brought storytellers a different tool for fulfilling their ideas, as well as for understanding particular characters or events that they are interested in. Nevertheless, I disagree with the argument that there is any *correct* ways for telling stories. Both the more “responsible” way, or the conventional cinematic storytelling, and the more “democratic” way, the navigation style of interactive storytelling, shall coexist in the future, similarly to how the Internet coexists with newspaper, television, and all the other media platforms. (And, we cannot deny the fact that a story that is complete and self-contained tends to be especially fascinating as well.) What’s important is, how we may leverage the computational technology to advance both storytelling forms, in order to help people to tell more stories that are inspiring, encouraging, and beneficial to the human society as well as our ecological environment [45, 46].

VII Related Work

In this chapter, I introduce the related work categorized in the following genres: commonsense computing, interactive storytelling, story representation and narrative theory, browsing interface for media collection, video retrieval, and automated video editing.

7.1 Commonsense Computing

Commonsense computing allows the concept-expansion functionality and tremendously benefits us in the machine-understanding process of the natural language questions and answers. The biggest knowledge base of common sense in the world is the CYC project [39]. Lenat started the CYC project with his colleagues in the 1980's. They use formal logic to represent the commonsense knowledge, and hire engineers to input the data in the last two decades. The Storied Navigation does not utilize this knowledge base because the formal logic it uses does not support analogy, which makes it difficult to perform the “ways to think” proposed in Chapter 2.

In 2000, Push Singh at the MIT Media Laboratory started the Open Mind Common Sense project (OMCS) [3], which is a website that asks internet users all over the world to input the common sense that they have, using the English language. Today, it has approximately 800,000 pieces of commonsense knowledge such as “A car has four wheels.” Using natural language processing techniques, the English sentences in OMCS were parsed into different semantic networks, including ConceptNet [2], LifeNet, and StoryNet [8], based on which dozens of applications have been developed [6, 8, 11]. ConceptNet is the largest semantic network based on OMCS. It is composed of 300,000 concepts and 1.6 million edges of 20 link types (e.g. “Effect Of”, “Location Of”, “Used For”, etc). It is the major commonsense knowledge source of Storied Navigation, since it is suitable for performing affect sensing for free-text input, as well as analogy [2, 14, 15].

7.2 Interactive Storytelling

7.2.1 Media Collection-Based Storytelling

Proposed by Davenport and Murtaugh in the late 90's, automatist storyteller systems (e.g., Dexter, ConTour [29, 30]) can be viewed as ancestors of Storied Navigation. These systems originate from the idea of collection-based viewing experience, where, in comparison with traditional films, no “final-cut” is made since the stories are created simultaneously as the viewing activity takes place.

Both ConTour and Dexter, similarly to Storied Navigation, build up graph relations of sequences and keywords and perform spreading activation over these graphs as the stories progress. Nevertheless, they are different from Storied Navigation in that 1) they are not using commonsense reasoning and are based on keywords, and, accordingly 2) the users cannot type in natural language sentences as story segments that they think of, such that the flow of storytelling cannot be as fluent and intuitive, and 3) the users lack the guidance of the questions and the answers, so it is relatively harder for them both to understand the material and to make the story progress.

Mindful Documentary [8] and our Storied Navigation both use commonsense reasoning technique to provide suggestions that are potentially helpful to the users. The difference is that the two processes are of two different stages in video storytelling activities. Mindful Documentary gives recommendation of what to film during the shooting process, which benefits users to collect materials that are of enough variety (since it is a crucial criterion to documentary shooting). Storied Navigation, on the other hand, tries to give inspiration on the progressions of stories, given a repository of existing video sequences.

Kevin Brooks's Agent Stories [13] is a system that helps users in the same stage of storytelling as Storied Navigation – editing and the construction of the story structure. It also provides story features like “conflict” or “negotiation”, but it does not use commonsense reasoning techniques to help users either to find the materials that they're thinking of, or to prompt questions to suggest possible progressions with a high generality of the stories.

Textable Movie is an interactive program which plays videos instantly as the user types in free text. The experience of making video stories by inputting story descriptions is shared by Textable Movie and Storied Navigation. What's different is that Storied Navigation gives multiple options based on commonsense reasoning while Textable Movie playbacks the videos instantly based on keyword matching. So Storied Navigation gives flexibility on the stories' progression, while Textable Movie gives instant feedback of how the result video looks.

Vista [75] is a system that allows users to search aural storytelling performances from a video database. The users can search by title, topic, performer's name, etc., select the video that they want to play, view the performance via a streaming media player, and simultaneously view the narrative transcript of the stories. They can also interact with the system about the narrative of the performers' stories by chatting or asking questions about themes, motifs, characters, and story content. It is similar to Storied Navigation: Both of them take natural language input and return videos in regards.

The differences between Vista and Storied Navigation are: 1) Vista does not provide mechanisms that facilitate annotation of users' own videos to create their own memories 2) Vista does plain search as conventional search systems do, and does not give inspirations to the users of what story paths might possibly be there, because there are too few "ways to think", only matching the things described in the questions and the answers 3) it does not leverage commonsense technology to analyze emotions, so it cannot perform story analogy.

Bangsø et al [76] proposed a Bayesian network framework that supports non-linear storytelling. More specifically, users can jump back and forth between different stages of a pre-defined story, based on the correlations between the stages carefully designed by the system developers. It a storytelling system different from the aforementioned approaches that, the users cannot add new story elements such as photos or video files to increase more possibilities of navigation paths. However, it is similar to Storied Navigation in terms of the interaction of the user than the next category that I am going to introduce, the character-based storytelling scenarios.

7.2.2 Character-Based Storytelling

Works in the area of character-based storytelling focus on the narrative experience the players who control virtual characters in a game environment have [48, 49, 50, 51, 52, 53, 54, 55, 56], including the notable Façade [48] and IDTension [49] systems. Façade is a widely known system for its strong functionalities facilitating character-based storytelling. Users can use natural language text to converse with the virtual characters in the story world, as well as to wonder around, to interact with objects in the 3D environment, etc. It supports free text dialogue input, and according to the author it performs better understanding of what the players possibly mean by the input text than keyword search. Nevertheless, the story world is a confined domain, so its text input is presumably less general than that in our Storied Navigation usage scenario. The natural language processing function does not leverage commonsense knowledge, so it cannot perform analogy or spreading activation for emotion sensing. IDTension, on the other hand, tries to combine "interactivity" and "narrativity" in the storytelling experience by allowing users to act as the directors of a performance. However, it is not for general storytelling activities either. It provides a constrained set of actions that the virtual characters can perform according to the players' selections, which is different from the usage scenario in Storied Navigation, where arbitrary characters, emotions and other descriptions can be used as the input. There are also many other work that focus on particular aspects or problems in character-based storytelling, such as natural language processing [51, 52], planning for individual virtual characters [53, 54, 56], emotion [55], and so on.

Although these inventions and findings are related to this thesis project, they are all focusing on

the role-playing activity, which is largely different to a video-editing/browsing scheme. That is, since in Storied Navigation we are dealing with material corpuses that the users try to make one or more story paths with, the problem is how to recognize the storytellers' goals in terms of "determining how to organize the sequences" or "designing a story". On the contrary, research on character-based storytelling aims to find the storytellers' goal in terms of interactions with other characters or objects in the story world such as finding, hiding, beating, utilizing, etc, so the focuses in two contexts are different, and should be represented with different goal models.

If we are to choose one work that is closest to Storied Navigation from the character-based area, Georgia Tech's Ari Lamstein and Michael Mateas's Search-Based Drama Management is probably the one most worth mentioning. The system SBDM (search-based drama management) helps guide the player's experience in more open-ended (than other work) story by evaluating the aesthetics within the player's actions. It performs a tree-search for the possible storylines from a large story corpus.

Nevertheless, none of the work in this category produce "cinematic" experience in the viewers' minds, which is a feeling produced by viewing juxtaposed sequences taken at different locations or time. Before the cinema appeared, people had always seen the world in a continuous viewing experience. The back-to-back cutting from one sequence to another arouses a viewing experience that happens only after the cinema appeared [82]. The viewing experience in virtual storytelling environment is analogous to our viewing experience in the real world, because the time and space in the story worlds are continuous too. For Storied Navigation, its retained basic building blocks – sequences – and ways of composing these building blocks – juxtapositions or "cuts" – from its predecessor, traditional cinema, endow the users with the privilege of telling stories in a cinematic way, which is one of the big differences between Storied Navigation and character-based storytelling in a game environment too.

7.3 Applications of or Discussions about Story Representation & Narrative Theory

Most of the works in this field present philosophical or computational analysis of stories or narratives, many of which make use of terminology like *syuzhet*, *fabula*, *narratology*, etc, [59, 60, 62, 63, 65]. Again, the word *syuzhet* means the temporal order that the event and characters in a story are narrated, whereas *fabula* stands for the inter-relationships between characters, between events, and between characters and events. Among all the story representation work, however, I particularly agree with Roger Schank's theories [17, 18] which does not leverage the *syuzhet* and *fabula* theory, because his theory of indexing stories and the relationship between intelligence and storytelling still greatly influence the related research today, even though his work in the 1970's was one of the pioneering work in this field. In addition, Schank's work is the best one that

matches my personal experience and observation in filmmaking. Below I also introduce other related works, in order to give a general sense of how analogous or different Storied Navigation is in the field.

Cheong and Young proposed a technique for generating the sense of “suspense”, or questions to the players that they need to find out the answers later, in an interactive role-playing game environment [59]. Briefly speaking, the system generates suspense by taking the fabula of a story and determines the syuzhet by reconstructing the ordering of the sentences. It is an interesting work, and it is interesting to think about the possibility of applying similar technique into Storied Navigation, but right now their work applies to a library of hand-crafted causality chains between the events in a confined story world, and it is difficult to apply it to corpuses that collect arbitrary video fragments of everyday life if they are not linked in any way. More specifically, we need to specify another “way to think” that human mind processes when having a question in mind that needs to be resolved, in order to find the knowledge representation or story features that are used in this process, and to build the fabula. Whether suspense really intrigues the viewers – or interactive storytellers – in a video corpus, as opposed to a role-playing game environment, is an interesting question to investigate as well.

Cavazza and Pizzi presented a comparison between several narrative theories and discussed their respective appropriateness for applying in interactive storytelling, particularly role-playing games [63]. According to my discussion in section 6.1, I suppose syuzhet and fabula are not necessarily applicable to Storied Navigation, but I would like to see if there would be researchers doing similar comparison for documentary-video-corpus-based interactive storytelling facilitated by commonsense reasoning

In order to give a computational representation for narratives and show how scientist can process the represented data for media files such as photos or videos, Goguen and Harrell talked about narrative structure using semiotics, algebraic semantics, cognitive linguistics, etc [61]. On the other hand, Tuffield et al [62] proposed a different approach –ontology – for similar goal purposes. These works have different goals from Storied Navigation, but they do provide fundamental resources that are potentially usable for further development of Storied Navigation in the future such as understanding of the story descriptions.

Narrative Prose Generation [60] is a natural language generation system for written narration, i.e. story description in my thesis. This system gives very impressive results – much better than other previous approaches. Nevertheless, it has a different goal from Storied Navigation – it generates the surface text from a given story structure, not the story structure per se. Also, the story domain is confined (500 concepts and 300 relations for three little riding hood narratives).

Finally, Candan et al. tried to build a representation for everyday “experience” instead of videos [64]. Nevertheless, a piece of experience is represented as a string of small events, similarly to the “primitive representation” that a story description in my thesis is parsed into. The matching between story descriptions in Storied Navigation can be extended by referencing Candan’s work, which might give even better results for the edit-by-typing function in my system.

7.4 Interface Design for Browsing/Annotating Media Collection

Ossenbruggen and Hardman [71] proposed an interface for users to annotate the time information for media files by drawing lines. It is a nice interface and it gives researchers important thinking specifically about time, for whom neither the design of a good interface for browsing and annotating media files nor a background representation have been addressed. This work shares part of the goals with Storied Navigation – proposing new annotation scheme that enables more intuitive browsing activities, but it is different from Storied Navigation that it does not give recommendations that keep viewers in a continuous path, which grows and extends along with new selections according to different criteria.

Appan et al [72] presented a system that helps users to annotate videos stories by who, what, when, where information, and features proposed by Kevin Brooks [13] that are fundamentally similar to our story structural roles. It also shares similar goals to our Storied Navigation that it tries to help more storied annotation for media files by helping the recording, managing, and sharing these media files, as well as experience sharing in a continuous storied flow. It even incorporates geographical information as part of the annotation for videos too. However, users always need to annotate both the story text and the story features. The system does not make use of the story text to alleviate users’ burden, and no story analogy or ways to think can be performed in the system.

Zhang and Nunamaker [73] also proposed a system that shares similar goals to Storied Navigation – helping people to navigate a media corpus by applying techniques that to some extent understand what the viewers’ want (natural language understanding in this example), and it gives nice results in the question-answer scenario. Similarly to the previously mentioned systems, however, it does not apply commonsense technology, and it cannot analyze emotions within story descriptions, nor can it perform analogy between story descriptions.

Wolff et al. [74] combined gameplay and narrative techniques to enhance the experience of viewing galleries. Again, this system also shares similar the goal of facilitating navigation of a media corpus. What is different from Storied Navigation is, it does not possess strong natural language processing functionalities that understand free text stories. It also works in a confined

story domain, where all the story paths are pre-determined. So it does not return a set of juxtaposed sequences for a written sentence, like Storied Navigation does.

Finally, ARIA (“Annotation and Retrieval Integrated Agent”) is another that facilitates media browsing or annotation. Our idea of suggesting annotations for videos is originated from ARIA’s integration technique of retrieval and annotation for photos by using commonsense reasoning technology [6]. However, the representations of the media pieces’ semantics in Storied Navigation (video sequence rather than photo) is different from that in ARIA. The different “ways to think” are not incorporated in ARIA either.

7.5 Video Retrieval based on Automated Annotation

The area of video retrieval based on content analysis techniques that try to extract meaningful features from video signals is a very huge. The most substantial works in this content-based retrieval field can be found in the TREC Video Evaluation Community (TRECVID) [83]. In this thesis document, I selected two representative works. First, Qi et al [78] presented a method that automatically annotates videos by hierarchical clustering, EM algorithm and probability voting process. While it is fully automatic, the semantics it derived are limited concepts like “indoor”, “landscape”, etc., which hardly contain any “storied” content and is not very useful for the “ways to think” for juxtaposition recommendation. Presented less than a month ago, this work shows us that there is still a long way to go for computers to derive fruitful enough semantics by “watching” the videos.

On the other hand, Lai et al. [79] extracted the content from speech data for annotation. It is an interesting and useful way for determining the relevance between videos, and it may be nice to integrate this technique into Storied Navigation. The audio signals that it extracts are the “discourse” of the characters and the sound in the environment when the video is recorded. But the “narrative” of the video, which is the description of what happens in the scene, is depicted more in the visual signals and may not be mentioned in the discourse. In order to make use of each video sequence in the database in a storied way, therefore, a piece of annotation that describes who, what, when, where information, is still needed.

Although I only introduce two proposed methods, works in this area are mostly based on a same philosophy – trying to find the semantics or even “interestingness” from video and/or audio signals. One fundamental difference between these works and Storied Navigation is that, they tend to ignore the importance of the subjective feeling that video owners have and the value of manual annotation, when they claim how labor-intensive and time-consuming manual annotation is. Often, however, what storytellers need as their story elements are the characters, events, or objects that

they find particularly meaningful, and whether it is true for a video will depend on their personal life experience. If a person is not acquainted with the video owners or is not shared with their life experience, sometimes it is hard for him/her to annotate it on behalf of them because of the lack of understanding about the context or their feelings, not to mention the computers. What we should do, before the computers can not only clearly understand individual videos by “watching” them but also build the surrounded context by relating a large set of videos, I think, is to help the users to do the annotation as easily as possible, by leveraging the least complex things they can possibly understand – common sense.

7.6 Other Work about Video Semantics

In 1996, Frank Nack presented a very nice application, AUTEUR [84], in his PhD thesis that performs automated film editing based on film theory, Roger Schank's CD (Conceptual Dependency) representation, and the commonsense data in Cyc. It shares many similar aspects with Storied Navigation. They both support video editing activities, they both help users to deal with large amount of video data, they are both based on Schank's theory, and they both utilize commonsense data from a large knowledge base. The differences are: 1) AUTEUR functions on the *shot* level of videos, whereas Storied Navigation functions on the *sequence* level, 2) AUTEUR focuses on humorous stories, whereas Storied Navigation is not limited to any kinds of stories, 3) AUTEUR is designed to tell stories that have only one character, whereas Storied Navigation can accommodate stories that have arbitrary numbers of characters, and 4) AUTEUR uses CYC as the commonsense knowledge base, where knowledge is coded in formal logic that does not support analogy; whereas Storied Navigation uses ConceptNet, which is a semantic network that helps performing analogy.

Among other works related to video semantics understanding, Shaw and Davis presented an analysis of how people annotate videos and how they search [80]. According to their study, the users tend to focus on the appearance of the characters instead of the roles that they play or their names. Since the study focused on Japanese animation videos (“anime”), as opposed to the documentary style videos that I use as the examples in my observation study, we cannot be sure this study result is applicable to Storied Navigation. However, it serves as a good suggestion on how we may perform similar study for documentary or maybe home videos, to investigate how our system can be improved.

Salway and Graham presented a paper that they claim as the first work trying to analyze emotions in film by not looking at the signals but other textual resources, in this case “audio description” [81]. They collected over 500 feature films all over the world, and utilized the accompanied audio description to extract affect characteristics for the characters in each scene. Although this method

is not applicable for documentary style video because there will be no such audio description, it is their idea of extracting affect from textual information instead of audio signals is analogous to our approach.

7.7 Automated Video Editing

7.7.1 Automated Documentary Video Editing

Lynda Hardman and her colleagues presented a system, Vox Populi [69], which tries to take documentary video sequences and generate a cut or a path automatically. In the examples they showed, the videos are interview sequences where people are sitting in front of the camera and talking about their perspectives in politics. Certainly, interview is one kind of useful resource in documentary-style storytelling. The design of Vox Populi, however, is somehow limited to this kind of video, and may not be as helpful for storytelling using other kinds of videos. That is, the criterion for finding the next clip is poor: only the “support-contradict” feature. It does not take into account the characters, the interaction between characters, or the analogy between scenes. No any “ways to think” for choosing the next coming sequence is provided besides using “conflict” and “resolution” as two types of story structural roles. The domain for this technique is also limited – it has to be interviews for controversial topics

Most of the other works that perform documentary video editing automatically are designed particularly for news videos, including Xiao Wu’s work that detects repeated ones and removes them [70], and the work by Duygulu et al. [77] that finds logos, or the graphical icons shown next to the anchor-person in news reports, in the videos. Both of them follow chronology and do not generate juxtaposition according to other types of ordering, because they do not perform any “ways to think”. In addition, for raw videos captured from end users’ video cameras that have no logos, the method that analyzes the logos in the video is not applicable.

7.7.2 Video Summarization

Truong and Venkatesh presented a systematic review and classification for Video abstraction and gave a comprehensive introduction and critiques about the existing methods of video abstraction [67]. They described video summarization as “mainly designed to facilitate browsing of a video database.”, and “also help the user to navigate and interact with one single video sequence in a nonlinear manner analogous to the video editing storyboard.” Therefore, to some extent video summarization has the same goal as Storied Navigation. The major difference is that, video summarization tends to be one piece of video artifact generated automatically by a system, whereas Storied Navigation is in itself a process, and the navigators can interactively decide what the “summarized” versions of the video corpus is. Pfeiffer et al.’s definition of a video abstract

clearly separates video abstracts from Storied Navigation: “a sequence of still or moving images that presents the content of a video compactly without losing the original message” [24]. After all, there is no real *original* message in the world of Storied Navigation.

Traditionally, works that were done to achieve the goals like “video summarization” and “video skimming” [23, 24, 25, 26, 27, 28] try to enable computers to summarize the input video by analyzing the video and audio signals such as camera motion and the loudness of sound. Generally, the outcome of video abstraction is generated based on no semantic processing, so none of these techniques can be used to generate stories that express causality or the transition of characters’ emotions, not to mention helping users to understand the materials. Recent works try to come up with novel representations such as intermediate-level ones that incorporate both camera motion and characters’ identities [31], “multimedia ontology” [32], or narrative structure graph [66], but they are either still limited in certain application domain, or do not facilitate the process of collection-based storytelling using free-text, or recommendations based on the “ways to think”.

VIII Conclusion

“...So that really allows you to navigate through his mind, basically.” – A user of Storied Navigation.

All human beings are by nature excellent storytellers. By telling stories, people share their experience, learn new knowledge, ask for other people's help, answer other people's questions, warn others of danger, negotiate with others on certain problems, and join numerous other activities. With the high accessibility of digital camcorders today, these inherent human capabilities could be applied in a more extensive and influential way if they could tell stories using these videos. Why? Because video is a rich type of media that it conveys massive information in a short amount of time, and that the Internet enables everyone to collaborate with others remotely as well as to broadcast their ideas to the world. What the video storytellers need, is a tool that helps them to focus on their stories, that helps them to deal with the large amount of video data, and that is easy enough so that everybody will be able to use.

This thesis document proposes a solution to this problem – Storied Navigation. I started this thesis by proposing two main questions, 1) “How can a system help users to develop their story threads?” and 2) “How can a system help users to gain familiarity with the materials stored in the corpus?”. Then, in Chapter 2 I introduced a theory based on my personal experience in making a documentary film, *Life. Research.*, as well as the existing literature in the AI storytelling area, particularly Roger Schank's theory on story indices. Based on these thoughts, I proposed nine types of “ways to think” that human beings perform in their mind when making video stories, and a set of “story features” for representing a story computationally, including the characters, the relationships between these characters, the emotions of these characters, the location, the date, and the theme of the story. Using these story features, I designed a system using the available commonsense computing resources, ConceptNet, MontyLingua, and WordNet. The system has various functions for supporting browsing and storytelling activities based on a video corpus, including edit-by-typing, “What's Next?” and “Find Alternatives” recommendations, search functions with different criteria such as character, emotion, theme, location, date, story structural role, and the traditional one, keyword. The system assists annotation by deriving the characters' roles, primitive representation, the possible theme, location, and date, based on the input story description, and a corpus of 81 annotated videos is used to conduct two evaluation studies.

The studies are described in Chapter 5. The first study was to investigate whether the system helps the users to construct storylines, and two subjects participated in the study. From the interview with the first subject, I found seven characteristics of the designed system; whereas from

interviewing the second subject I came up with five hypotheses, and derived a direction for future development of the system. The second study presents the subjects' perspectives on how my system may be used as a browsing tool for an unfamiliar video corpus, as well as some numerical analysis of the usage of each individual function. Then, an analysis of the questionnaire filled by all the nine subjects is presented, followed by a series of discussion in Chapter 6, and the related literature in Chapter 7.

Generally speaking, the previous chapters give a full presentation of what I have thought about and how I thought about it, over the process of this thesis project. Even though some of the features may not work ideally according to the participants in the studies, the overall result seems encouraging and – to myself – exciting. The next steps of this project may be incorporating the system in real documentary film projects, inviting more novice users to make video blogs using this system online, creating browsing interfaces for large video collections such as the Media Lab video database or even YouTube, as well as pushing it to the next level in terms of computational capability by finding the ways to implement more “ways to think”, and by creating plan recognition models that can inference what the storytellers are thinking about – in terms of story structure, in particular – during the storytelling process.

At the end, I would like to tell a little story to acknowledge a subject who participated in my study. I chatted with this subject after the study session, and she encouraged me to continue working on this topic, when we talked about my career in the future. The reason why I would like to acknowledge this person here is not simply because of her encouragement, but also something insightful in her words, which is directly related to this research project. I told her that I actually enjoy telling stories much more than making storytelling software for people to use, so I was not certain whether I should go with the direction of storytelling, instead of technology. Her reaction was, there is no conflict between them. Initially I found it somewhat cliché, because I had heard this type of argument for many times. But as she moved on, (I don't remember perfectly what she really said, but what it was like this), she said “There's too much violence and bad things in the world. In our TV, on the internet...you can find those things everywhere, because it is too easy to make something violent and show it to people. But it's so hard to make the beauty. Your film of Hugh is beautiful, but there are a lot of people who want to make such beautiful films about their friends and families too. And they cannot, because it's too hard...” Her words answered my question about the *real* motivation of continuing this project – not the practical motivation in terms of a technical thesis, but the motivation with regard to my own life. Most technologies are invented because the inventors wish the human race to have a more enjoyable way of living, a better social society, or a better environment. I invented Storied Navigation because, I guess, I wish the people who have the chance to make a story about someone like Hugh or someone they love, can have the tool to make it.

IX Reference

1. C. Fellbaum (Ed). WordNet: An Electronic Lexical Database. MIT Press. 1998.
2. H. Liu and P. Singh. ConceptNet: a practical commonsense reasoning toolkit. *BT Technology Journal*, 22(4):211-226. 2004.
3. P. Singh, T. Lin, E. T. Mueller, G. Lim, T. Perkins, and W. Li Zhu. Open Mind Common Sense: Knowledge acquisition from the general public. *Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*. Lecture Notes in Computer Science. Heidelberg: Springer-Verlag. 2002.
4. H. Liu. MontyLingua: Commonsense-informed natural language understanding tools. 2003. Available at: <http://web.media.mit.edu/~hugo/montylingua/>
5. Roballet. <http://weblogs.media.mit.edu/roballet/>
6. H. Lieberman and H. Liu. Adaptive linking between text and photos using common sense reasoning. 2002.
7. C. Vaucelle and G. Davenport. An open-ended tool to compose movies for cross-cultural digital storytelling: Textable Movie. *Proceedings of ICHIM 04 'Digital Culture & Heritage'*. 2004.
8. B. Barry. *Mindful Documentary*. PhD Thesis, MIT Media Lab. 2005.
9. Y.-T. Shen. Knowledge-based-aided video editing, Technical Report, Available at <http://graphics.csie.ntu.edu.tw/~edwards/VideoEditing.pdf>. 2005.
10. M. Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. Perennial, 1991.
11. H. Lieberman, H. Liu, P. Singh, and B. Barry. Beating Common Sense Into Interactive Applications. *Artificial Intelligence Magazine*, 25(4), 63-76. AAAI Press. 2004.
12. M. L. Murtaugh, *The Automatist Storytelling System: Putting the Editor's Knowledge in Software*, Master Thesis, MIT Media Lab. 1996.
13. K. M. Brooks. *Agent Stories: Authoring Computational Cinematic Stories*. PhD Thesis, MIT Media Lab. 1999.
14. M. Minsky. *The Society of Mind*. Simon & Schuster, 1988.
15. M. Minsky. *The Emotion Machine*. 2006.
16. R. G. Evans. *LogBoy Meets FilterGirl: A Toolkit for Multivariant Movies*. Master Thesis, MIT Media Lab. 1994.
17. R. Schank. *Tell Me a Story: A New Look at Real and Artificial Intelligence*. 1991.
18. R. Schank and R. Abelson. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. 1977.
19. R. Schank and C. K. Riesbeck. *Inside Computer Understanding*. 1981.

20. L. Morgenroth. *Movies Talkies Thinkies: An Experimental form of Interactive Cinema*. Master Thesis, MIT Media Lab. 1995.
21. YouTube™ <http://www.youtube.com>
22. D. Bordwell. *Narration in the Fiction Film*. University of Wisconsin Press. 1985.
23. A.A. Wolf and L. Wolf. Video de-abstraction or how to save money on your wedding video, in *Proceedings of Applications of Computer Vision, 2002*. (WACV 2002), 264- 268. 2002
24. Pfeiffer, S., Lienhart, R., Fischer, Stephan and Effelsberg, W. Abstracting Digital Movies Automatically. In *Journal of Visual Communication and Image Representation*. Vol. 7, No. 4, 345-353. 1996.
25. R. Lienhart. Abstracting home video automatically. *ACM Multimedia* (2) 1999: 37-40. 1999.
26. X.-S. Hua, L. Lu, and H. J. Zhang. AVE: automated home video editing, in *Proceedings of ACM Multimedia 2003*, 490-497. 2003.
27. Y.F. Ma, L. Lu, HJ Zhang and M. Li A User Attention Model for Video Summarization, in *Proc. of ACM MM 2002*, France, pp.533-42. 2002.
28. Z. Rasheed, Y. Sheikh, and M Shah. On the use of computable features for film classification, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 1, No.11,2003.
29. G. Davenport and M. Murtaugh. Automatist storyteller systems and the shifting sands of story, in *IBM Systems Journal*, Vol. 36, No. 3.
30. JBW: A Random Walk through the 20th Century <http://ic.media.mit.edu/projects/JBW>
31. X. Tong, Q. Liu, L. Duan, H. Lu, C. Xu, Q. Tian. A Unified Framework for Semantic Shot Representation of Sports Video, in *Proc of ACM MIR 2005*, pp127-134. 2005
32. M. Bertini, A.D. Bimbo, C. Torniai. Automatic Annotation and Semantic Retrieval of Video Sequences using Multimedia Ontologies. In *Proc of ACM MM 2006*, pp679-682. 2006.
33. G. Davenport. Interactive Multimedia on a Single Screen Display. In *Videotechnology Technical Session, Current Applications of Videotechnology in Computer Graphics Systems*, National Computer Graphics Association Conference, March 22, 1988.
34. H. Liu and P. Singh Commonsense Reasoning in and over Natural Language, *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES'2004)*. Wellington, New Zealand. September 22-24. Lecture Notes in Artificial Intelligence, Springer 2004, 2004.
35. M. Minsky. Commonsense-Based Interfaces. In *Communications of the ACM*, Volume 43, Issue 8, 2000.
36. H. Chung. GlobalMind - Bridging the Gap Between Different Cultures and Languages with Common-sense Computing. Master Thesis, MIT Media Lab, 2006.
37. Lenat, D.B. 1995. CYC: A large-scale investment in knowledge infrastructure. In *Communications of the ACM*. 38(11): 33-38, 1995.

38. Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. eds. *Embodied Conversational Agents.*: MIT Press. 2000.
39. Fano, A. and Kurth, S. W. 2003. Personal Choice Point: Helping users visualize what it means to buy a BMW. *Proc of IUI 2003*: 46-52.
40. Fleischman, M. and Hovy, E. 2003. Recommendations without User Preferences: A Natural Language Processing Approach. *Proc of IUI 2003*: 242-244. 2003.
41. Liu, H., Lieberman, H., and Selker, T. 2003. A Model of Textual Affect Sensing using Real-World Knowledge. *Proc of IUI 2003*. 2003..
42. Liu, H. and Maes, P. 2005. InterestMap: Harvesting Social Network Profiles for Recommendations. *Proc. of the Beyond Personalization 2005 Workshop*. 2005
43. Resnick, P. and Varian, H. R. 1997. Recommender Systems. *Communications of the ACM*, 40(3): 56–58. 1997
44. Mehrabian, A. *Manual for a comprehensive system of measures of emotional states: The PAD Model.* (Available from Albert Mehrabian, 1130 Alta Mesa Road, Monterey, CA, USA 93940). 1998.
45. S. Bognar and J. Reichert. *A Lion in the House*. Documentary film. A Lion in the House, LLC. 2006.
46. *An Inconvenient Truth*. Documentary film. Paramount Classics. 2006.
47. G. Davenport. *Sharing Video Memory: Goals, Strategies and Technology* (Short Working Paper). Cambridge, MA: MIT Media Lab. 2005.
48. M. Mateas and A. Stern. *Façade: An Experiment in Building a Fully-Realized Interactive Drama*. In *Proc. of Game Developer's Conference 2003*, San Jose, CA. 2003.
49. N. Szilas. *IDtension: a narrative engine for Interactive Drama*. In *Proc. of Technologies for Interactive Digital Storytelling and Entertainment (TIDSE'03)*. Fraunhofer IRB Verlag, 2003, 187-203.
50. M. O. Riedl. *Narrative generation: balancing plot and character*. PhD Thesis. North Carolina State University. 2004.
51. S.J. Mead, M. Cavazza and F. Charles. *Influential Words: Natural Language in Interactive Storytelling*. In *Proc of 10th International Conference on Human-Computer Interaction*, Crete, Greece. 2003.
52. M. Mateas and A. Stern. *Natural Language Understanding in Façade: Surface-text Processing* In *Proc of Technologies for Interactive Digital Storytelling and Entertainment (TIDSE)*, Darmstadt, Germany, June 2004

53. Y. Cai , C. Miao , A.-H. Tan , Z. Shen. Fuzzy cognitive goal net for interactive storytelling plot design. In *Proc of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, Hollywood, California, June 14-16, 2006.
54. L. Motta Barros and S. R. Musse. Planning Algorithms for Interactive Storytelling. In *Computers in Entertainment (CIE)*, vol 5, issue 1, 2007
55. D. Pizzi, F. Charles, J.-L. Lugin and M. Cavazza. To appear in *Proc of Interactive Storytelling with Literary Feelings*. ACII, Lisbon, Portugal, September 2007.
56. R. M. Young. Cognitive and Computational Models in Interactive Narratives. In *Cognitive Systems: Human Cognitive Models in Systems Design*, Chris Forsythe, Michael L. Bernard & Timothy E. Goldsmith, editors, Lawrence Erlbaum.
57. J. H. Murry. *Hamlet on the Holodeck*. MIT Press. 2000.
58. A. Lamstein and M. Mateas. Search-Based Drama Management. In D. Fu, S. Henke, & J. Orkin (Eds.) *Challenges in Game Artificial Intelligence: Papers from the AAAI Workshop* (Technical Report WS-04-04). Menlo Park, CA. 103-107.
59. Y.-G. Cheong and R. M. Young. A Computational Model of Narrative Generation for Suspense. In *Proc of the AAAI 2006 Workshop on Computational Aesthetics*, 2006.
60. C. B. Callaway and J. C. Lester. In *Artificial Intelligence*, 139(2): 213-252, August 2002.
61. J. Goguen and F. Harrell. Foundations for Active Multimedia Narrative: Semiotic spaces and structural blending. In *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*. 2004
62. M. M. Tuffield, D. E. Millard and N. R. Shadbolt. Ontological Approaches to Modeling Narrative. In *Proc. of 2nd AKT DTA Symposium*, AKT, Aberdeen University. 2006.
63. M. Cavazza and D. Pizzi. Narratology for Interactive Storytelling: a Critical Introduction. In *Proc of 3rd International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE)*, Darmstadt, Germany, December 2006.
64. K. S. Candan, M. E. Donderler, J. Ramamoorthy, J. W. Kim. Clustering and indexing of experience sequences for popularity-driven recommendations. In *Proc of 2006 ACM Workshop on Capture, Archival and Retrieval of Personal Experiences (CARPE)*. 2006
65. V. Propp. *Morphology of the Folktale*. Texas Press. 1968.

66. B. Jung, J. Song and Y. Lee. A Narrative Based Abstraction Framework for Story-Oriented Video. In *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol.3, No.2, Article 11, May 2007
67. B. T. Truong and S. Venkatesh. Video abstraction: A systematic review and classification. In *ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMCCAP)*, 3(1), Jan 2007.
68. X. Shao, C. Xu, N. C. Maddage, Q. Tian, M. X. Kankanhalli, and J. S. Jin. Automatic summarization of music videos. In *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP) Vol. 2 , Issue 2*, May 2006.
69. S. Bocconi, F. Nack, and L. Hardman. Vox Populi: a tool for automatically generating video documentaries. In *Proc. of the sixteenth ACM conference on Hypertext and hypermedia 2005*, Salzburg, Austria, September 2005.
70. X. Wu. Threading Stories and Generating Topic Structures in News Videos across Different Sources. In *Proc of 13th Annual ACM International Conference of Multimedia (ACM MM'05)*, Singapore, Nov. 2005.
71. J. v. Ossenbruggen and L. Hardman. Semantic Timeline Interfaces for Annotated Multimedia Assets. In *Proc. of The 2nd European. Workshop on the Integration of Knowledge,. Semantic and Digital Media Technologies. (EWIMT)*. 2005.
72. P. Appan, H. Sunaram and D. Birchfield. Communicating everyday experiences. In *Proc of the 1st ACM workshop on Story Representation, Mechanism and Context*. 2004.
73. D. Zhang and J.F. Nunamaker. A Natural Language Approach to Content-Based Video Indexing and Retrieval for Interactive e-Learning. In *IEEE Transactions on Publication*, Vol 6, Issue 3, 2004.
74. A. Wolff, P. Mulholland and Z. Zdrahal. Combining Gameplay and Narrative Techniques to Enhance the User Experience of Viewing Galleries. In *ACM Computers in Entertainment (CIE) Volume 5 , Issue 1* (January 2007)
75. Elizabeth Figa and Paul Tarau, The VISTA Project: An Agent Architecture for Virtual Interactive Storytelling, in *Proc of TIDSE 2003*, Darmstadt, Germany, March 2003.
76. O. Bangsø, O. G. Jensen, F. V. Jensen, P. B. Andersen, and T. Kocka. Non-linear interactive storytelling using object-oriented Bayesian networks. In *Proc of International Conference on Computer Games: Artificial Intelligence, Design and Education*, 2004.

77. P. Duygulu, J.-Y. Pan, and D. A. Forsyth. Towards auto-documentary: tracking the evolution of news stories. In *Proc. of ACM Multimedia 2004*, New York City, NY, pp. 820-827, October 10-16, 2004
78. G.-J. Qi, X.-S. Hua, Y. Song, J. Tang, and H.-J. Zhang. Transductive Inference with Hierarchical Clustering for Video Annotation, In *Proc of IEEE International Conference on Multimedia & Expo (ICME)*, Beijing, China, July, 2007.
79. W. Lai, X.-S. Hua, and W.-Y. Ma. Towards content-based relevance ranking for video search. In *Proc of ACM Multimedia 2006*, 627-630, 2006.
80. R. Shaw and M. Davis. Toward emergent representations for video. In *Proc of ACM Multimedia 2005*, 431-434. 2005.
81. A. Salway and M. Graham. Extracting Information about Emotions in Films. In *Proc. of the eleventh ACM international conference on Multimedia*, 299-302, 2003.
82. W. Murch. *In the Blink of an Eye*. 2nd vers !"#\$. Silman-James Press, 2001.
83. TREC Video Evaluation Community (TRECVID). <http://www-nlpir.nist.gov/projects/trecvid>
84. F. Nack. *AUTEUR: The Application of Video Semantics and Theme Representation for Automated Film Editing*. PhD Thesis. Lancaster University. 1996.
85. R. Kearney. *On Stories*. Routledge. 2002
86. A. Rosenthal. *Writing, Directing, and Producing Documentary Films and Videos*. Southern Illinois University, 3rd Edition, 2002.