Raconteur: Intelligent Assistance for Conversational Storytelling with Media Libraries

BY

PEI-YU (PEGGY) CHI

MASTER OF SCIENCE IN COMPUTER SCIENCE
NATIONAL TAIWAN UNIVERSITY, 2008

BACHELOR OF BUSINESS ADMINISTRATION IN INFORMATION MANAGEMENT
NATIONAL TAIWAN UNIVERSITY, 2006

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 MEDIA ARTS AND STUDIES

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2010

Signature of Author ______________________________
Program in Media Arts and Sciences
September 10, 2010

Certified by ______________________________
Dr. Henry Lieberman
Research Scientist
MIT Media Laboratory
Thesis Advisor

Accepted by ______________________________
Prof. Pattie Maes
Associate Academic Head
Program in Media Arts and Sciences

Raconteur: Intelligent Assistance for Conversational Storytelling with Media Libraries

by

Pei-Yu (Peggy) Chi

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 Media Arts and Sciences
at the Massachusetts Institute of Technology

September 2010

ABSTRACT

People who are not professional storytellers sometimes have difficulty putting together a coherent and engaging story, even when it is about their own experiences. However, consider putting the same person in a conversation with a sympathetic, interested and questioning listener, suddenly the story comes alive. There's something about the situation of being in a conversation that encourages people to stay on topic, make coherent points, and make the story interesting for a listener.

Raconteur is a system for conversational storytelling between a storyteller and a viewer. It provides intelligent assistance in illustrating a life story with photos and videos from a personal media library. Raconteur performs natural language processing on a text chat between two users and recommends appropriate media items from the annotated library, each file with one or a few sentences in unrestricted English. A large commonsense knowledge base and a novel commonsense inference technique are used to understand event relations and determine narration similarity using concept vector computation that goes beyond keyword matching or word co-occurrence based techniques. Furthermore, by identifying larger scale story patterns such as problem and resolution or expectation violation, it assists users in continuing the chatted story coherently. A small user study shows that people find Raconteur's suggestions helpful in real-time storytelling and its interaction design engaging to explore stories together.

Thesis supervisor: Henry Lieberman
Title: Research Scientist of MIT Media Laboratory

Raconteur: Intelligent Assistance for Conversational Storytelling with Media Libraries

by

Pei-Yu (Peggy) Chi

September 2010

Thesis Reader

Dr. Nick Montfort
Associate Professor of Digital Media
Program in Writing and Humanistic Studies, MIT

Raconteur: Intelligent Assistance for Conversational Storytelling with Media Libraries

by

Pei-Yu (Peggy) Chi

September 2010

Thesis Reader ____________________________________

Dr. Fox Harrell
Associate Professor of Digital Media
Program in Writing and Humanistic Studies, MIT
Program in Comparative Media Studies, MIT
MIT Computer Science and Artificial Intelligence Lab

Acknowledgement

This thesis could not have been accomplished without the tremendous supports from the following great researchers and friends:

My advisor Henry Lieberman is like a father on my research path: he opened a brand new perspective on AI, HCI, and storytelling to me, and introduced the important concepts and attitudes about research, career planning, and life. Moreover, he greatly inspired me to shape this research when we "chatted" about many "travel stories" as we are both frequent travelers and went on a three-week trip to Asia for the conference IUI 2010 and sponsor visits this February. I'm deeply grateful to become Henry's student at the Media Lab, where I spent such pleasant two years with several exciting research projects.

A very important basis of this thesis comes from literature and narrative studies. I'd like to express my deepest gratitude to my readers who brought me to this inspiring, graceful field, introduced the significant resource, and provided their insightful comments: Prof. Nick Montfort, who offered a wonderful class of interactive narrative with the support of his excellent research, open-minded attitudes, and logical ways of thinking, and Prof. Fox Harrell, who involves in both AI and narrative research with his innovative approaches and wide knowledge on different fields. Because of them, I've been enchanted to the world of computational narratives, and will continue my adventure.

The technical aspect of my thesis could have only been possible with the enormous efforts and supports from the Software Agents group. I sincerely thank my colleagues Catherine Havasi, Rob Speer, Dustin Smith, Ken Arnold, Jayant Krishnamurthy, Jason Alonso, Ian Eslick, and many others who participated in our group. Their continuous, focused, and high quality research on commonsense reasoning techniques and applications has gradually made the dream of AI come true.

The encouragement from the lab sponsors and visitors are gratefully acknowledged, especially: ITRI (Industrial Technology Research Institute) from Taiwan for their fellowship support of my studies in year 2010 and the workshop opportunity together with Henry and Rob, Hallmark for their high interests and many valuable discussions, and InfoCast for their interests and hosting us in Hong Kong. I'd also like to thank the participants of the Media Lab sponsor weeks during these two years, and all the reviewers and participants in the conferences IUI 2010, ELO 2010, and CHI 2009 for their feedback on refinement over these research iterations. In addition, I highly appreciate the help from the director of our academic program administration Linda Peterson and the administration assistant Aaron Solle, who have kindly helped me out several times even when I was desperate in certain situations. I truly appreciate their assistance and cherish this degree even more.

I thank my friends from the lab, school of architecture, and CSAIL: It was such pleasant experience to work with the talented women Keywon Chung (Tangible Media group), Carnaven Chiu (program of Design Computation and the SENSEable City lab), and Xiao Xiao (Tangible Media group) as the team "ChiaoChiuChiChung" - our final course project that was published in CHI 2009 initiated my interest in the problem of digital media appreciation and storytelling. Later with the talented "mother" Angela Chang (Personal Robots group) I had a chance to explore narrative theories, publish and present our paper about children's interactive narrative design at the conference ELO 2010 together. It was enjoyable to work and discuss various research topics with Chih-chao Chuang (Smart Cities group), who always passionately shared his great talent on design and architecture. I believe the grad student life with countless dinners and working nights with the smart and hard workers Carl Yu (program of Design Computation and the Design Lab) and Wei Dong (Future of the Opera) will be one of the things I miss much. Tsung-hsiang Chang (CSAIL), with whom I prepared and moved to Boston from Taiwan together, gave me ideas of network infrastructure design to realize the Raconteur's chat scenario. Finally, it was inspiring to interact with the lab students Agnes Chang, Nan-wei Gung, Ahmed Kirmani, Sophia Yuditskaya, Michael Lin, Aithne Pao, and the alumni Dori Lin, Edward Shen, and Chao-chi Chang during these years.

From the bottom of my heart I'd like to give my special thanks to our Media Lab Taiwanese "leaders" Wu-hsi Li, who generously offered his enormous helps and care with my application, interviews, grad student life, and all the way to my thesis, and Jackie Lee, who invited me to join the Nightmarket workshop from year 2006, which started my journey to the Media Lab. My longtime mentors Prof. Hao-hua Chu, Prof. Robin Bing-yu Chen, and Prof. Jane Yung-jen Hsu at National Taiwan University have been guided and influenced me.

I wish to extend my warmest thanks to friends at MIT, Boston, and Taiwan – there are too many people for whom I have great regard. Because of all the precious memories we have, I could personally experience and observe many wonderful life stories. Here I'd like to particularly single out those who supported this study in various ways: Wen-hsuan Lee, Hsiang-chieh Lee, Cheng-wei Cheng, I-hua Chen, Chien-jen Lai, Michelle Bau, Ingrid Bau, Jacqueline Lee, Peggy Quinn, Grace Shih-Ling Hsu, Yen Tien, Mike Chen, Cathy Hsu, Yen-ju Chen, Teressa Kuo, and Tang-yu Cheng.

Finally and most importantly, this thesis is dedicated to my dearest family: my father, my mother, my sister, my brother in law, and Takumi Yamamoto, who are always there to have a "conversation" and a laugh with me. Their love, care, inspiration, patience, and encouragement, have made all of this possible and positively supported me moving forward.

Because of these wonderful contributors,
this thesis has opened another new chapter of my life.

Contents

Acknowledgement 9

Contents 13

List of Figures 17

List of Tables 19

Chapter 1 Introduction 21

1.1 Motivation 21

1.2 Problem Definition and Proposed Solution 23

1.3 Scenario of Raconteur: From Chat to Stories 24

1.4 Design Challenges and Contribution 28

1.5 Thesis Organization 29

Chapter 2 Background 31

2.1 Narrative Theory 31

2.1.1 Narrative and Its Components 32

2.1.2 Interactive Narrative 33

2.1.3 Moving Toward Digital Personal Narrative 35

2.2 Conversational Storytelling 36

2.3 Story Analogy and Patterns 38

2.4 Commonsense Reasoning 39

2.4.1 Storytelling and Common Sense 39

2.4.2 OMCS Knowledge Base 40

2.4.3 AnalogySpace Inference Techniques 42

2.5 Summary 44

Chapter 3 Formative User Studies 45
3.1 Study #1: Chat and Digital Media 45
3.1.1 The User Interface and Experimental Setup 46
3.1.2 Results and Discussion 47
3.2 Study #2: Story Composition and Patterns 48
3.2.1 Story Collection 48
3.2.2 The User Interface and Experimental Setup 50
3.2.3 Results and Discussion 51
3.3 Summary 52
Chapter 4 System Design and Implementation 53
4.1 Raconteur Structure 53
4.2 Resources of Multimedia and Assumptions 55
4.3 Narration Processing and Representation 56
4.3.1 Natural Language Processing 57
4.3.2 Non-story-world Clause Removal 59
4.4 Analogical Inference 60
4.4.1 Building Concept Vectors 60
4.4.2 Associating Similar Elements 61
4.4.3 Story Pattern Reasoning by Making Analogies 63
4.4.4 Finding Story Paths 70
4.5 Story Developer 71
4.6 Implementation 71
Chapter 5 Chatting through Raconteur 73
5.1 Preparing for Chat 73
5.2 Storytelling as Easy as Chat 76
5.3 Editing with Suggestions 77
5.4 Observing Story Patterns 78
Chapter 6 Evaluation and Discussion 79
6.1 User Study Goals 79

6.2 Participants and Material 79
6.3 Procedure 81
6.4 Measurement 82
6.5 Results and Discussion 83
6.5.1 Quantitative Results 83
6.5.2 Qualitative Results 86
Chapter 7 Related work 93
7.1 Chat, Collaboration, and Multimedia 93
7.2 Media Composition using Natural Language 94
7.3 Dialogue Systems and Models 95
Chapter 8 Conclusion and Future Work 97
References 99

List of Figures

Fig. 1-1. Screenshots from Facebook (top) and Picasa (bottom): online users chat through friends' shared personal media on social network websites. 23

Fig. 1-2. Raconteur user interface, including: a) a chat box where storytellers can chat with a friend in text, see the matched media elements, and edit to enhance his story, b) a preview window to see the photos and videos with captions, and c) Raconteur's suggestion panel for observing the story patterns and the multimedia repository. 25

Fig. 2-1. Transforming a knowledge base into matrixes by SVD [Havasi 2009] 43

Fig. 3-1. User interface for the formative study of chat, where storytellers can: a) see the multimedia repository, b) chat with a story viewer, c) compose the media elements, and d) observe the basic search result. 46

Fig. 3-2. User interface for the formative study of story patterns: (a) the upper part presents the raw material of the unorganized collection (a2), and provides the editing interface for users to decide the story goal (a1) and the sequence of scenes (a3); (b) the lower part shows the sets of analogous story elements in a pattern that matches the story goal. 51

Fig. 4-1. Raconteur system structure 54

Fig. 4-2. Algorithm of matching two media elements in the simplest way 62

Fig. 4-3. Algorithm of finding a collection of problems and resolutions............... 65

Fig. 4-4. Algorithm of finding collections bringing different emotions 68

Fig. 5-1. User uploads media elements to an online album and annotates with captions. 74

Fig. 5-2. Raconteur's login interface: by reading the online album information, the system allows a user to choose a story to chat. 74

Fig. 5-3. The initial state of the Raconteur interface with a chat box, preview box, and the raw set of material fetched from the online album............ 75

Fig. 5-4. User can preview the media elements by mouse hover. 75

Fig. 5-5. Chat box design of the teller's side (top) and viewer's side (bottom), each including a) an input box, b) a list of chat messages and files, and edit areas for the storyteller, and c) Raconteur's found elements that match the latest narration. 76

Fig. 5-6. User can directly chat on any media element in the system. For example, the story viewer responds to the teller by posting a question specific to one of the edited photos. 77

Fig. 5-7. Edit elements to enhance narrations by drag-and-drop. 77

Fig. 5-8. Raconteur's suggestion panel of story patterns 78

List of Tables

Table 2-1. The set of the defined relations that connect concepts in ConceptNet 4 41

Table 2-2. Examples of similar concepts found by AnalogySpace 43

Table 3-3. Multiple story units with similar patterns. The number indicates the elements in the same set of patterns. 50

Table 4-1. Examples of stories behind a still photo and a short video clip taken from a trip. 56

Table 4-2. Selected simplified POS tagset in NLTK 58

Table 4-3. The processed sentence structure as a result of natural language processing 59

Table 4-4. Lists of different types of non-story-world words and clauses in conversations 60

Table 4-5. Lists of matched examples to the pattern of problem and resolution 64

Table 4-6. Dot product results of the desire vector and exemplar vectors of concepts, where the negative value indicates people do not like that concept as much as other concepts with positive values 65

Table 4-7. Selected Raconteur's results from different repositories to the pattern of problem and resolution 66

Table 4-8. Lists of matched examples to the pattern of expectation violation 67

Table 4-9. Lists of examples to the pattern of emotions 69

Table 6-1. Details of participants' uploaded media sets for the study 81

Table 6-2. The facts about the five chat sessions 83

Table 6-3. The analysis of the numbers of the media elements being edited by storytellers 83

Table 6-4. Analysis of results from the questionnaire using Likert-5 scale, in which a 5 means "strongly agree" and 1 as "strongly disagree" 85

Chapter 1

Introduction

1.1 Motivation

Technology enables us to easily and quickly capture our daily life moments with photos and video clips, and share with others through the online social platforms such as Flickr, Picasa, Facebook, and YouTube. Usually, such a personal multimedia system is full of individual media elements that include various story events illustrating different points and subjects. However, when showing a large set of material to a friend, such as going on a one-week vacation, it is not easy to understand how the events can be presented coherently and meaningfully, and what the audience is interested to see. Most people, especially novice users, therefore choose to present the story by events in chronological order [Kirk *et al.* 2006]. They often do not pay attention to the "point" being made by showing a given scene and whether the content is engaging to the audience, making it difficult for the viewers to follow the story like a dreary slideshow. As a result, as capturing and sharing become more accessible to amateur users, increasing quantity of such content can be easily found on the Internet. Although there is software for automating categorization by locations or time-stamps, it is still challenging to create a coherent presentation that tells an entertaining story from a higher level. We believe that an intelligent interface that provides assistance in relating the concrete elements of the scene to the overall story intent and considering viewer's interests, will result in more effective story composition.

For a set of captured media material, if we put the users into a face-to-face conversation with a sympathetic, interested and questioning listener, suddenly *the stories come alive*. There's something about the situation of being in a conversation that encourages people to stay on a topic, make connected points, and tell an interesting story to a listener.

Conversational storytelling is one of the basic, familiar forms of human communication in our everyday life. It involves at least one speaker and one listener to continue creating stories together, making the storytelling process interactive and responsive. It is as easy as having casual conversation, but at the same time has a purpose to share life stories, which are usually composed of important narrative elements such as characters, events, and causal connections. Not only the narrator is responsible to make coherent and tailored statements, but also the listener needs to respond and acknowledge what is just said. In this way, the created stories are usually reportable and structural that help the story recipients understand the specific context and communicate better.

Because of the easy access to the Internet, conversations over digital media between online users become more and more common. We observe the trend of social media that average users not only share personal multimedia data, but also associate with contextual textual information such as adding captions or comments, changing album titles or filenames, making subtitles, etc., in order to communicate their intentions and opinions behind the media to friends or the public. At the same time, the audience usually responds with their comments or analogous personal stories (Fig. 1-1), and this motivates the authors to answer and tell more about the experience. In other words, people *chat about life stories through digital media* to know more about each other over the Internet. This user interaction provides the opportunity for intelligent systems to understand the story intent behind digital media elements based on human conversations, and further help users to create stories with interesting, connected points by suggesting fitting elements of high reportability.

Fig. 1-1. Screenshots from Facebook[1] (top) and Picasa[2] (bottom): online users chat through friends' shared personal media on social network websites.

1.2 Problem Definition and Proposed Solution

Storytelling in the digital world can be closer to the real life experiences for novice users. Instead of directly putting users into an unfamiliar situation of assembling individual media elements from scratch, we aim at helping them to focus on their high-level stories and to

---

[1] http://www.facebook.com/

[2] http://picasa.google.com/

communicate with the audience through an interactive process, as what they usually do in daily conversations.

In this thesis, we present Raconteur, a personal story editing system that helps users think about story development in multimedia material by enabling conversations with friends. The word "raconteur", by definition, is a person who is skilled in relating stories and anecdotes meaningfully. Raconteur enables a dialogue between the storyteller and the viewer to develop a story – the viewer posts a question, and the storyteller answers with story details. Raconteur presents analogous media elements with goals that match the user's intention, and suggests story units for the storyteller to continue. Using natural language processing, analogical inference, and Commonsense reasoning, Raconteur analyzes the multimedia items in a repository, each annotated with textual information, to find story patterns and paths.

For example, a user may present his story of a trip based on the surprising moments (such as viewing the city from a high tower, encountering famous art in a local park, etc.) or by culture shock (e.g. having difficulty to read the menu in a restaurant, trouble communicating with the bus driver, etc.), while each of these story paths may result in different experiences to the audience. When the user chats with a viewer, Raconteur processes the chat messages in real time, reasons about the story intentions and viewer's interests, finds relevant elements, and suggests the story sequences to support the teller's point.

1.3 Scenario of Raconteur: From Chat to Stories

This section presents a description of Raconteur's web-based user interface (shown in Fig. 1-2), and explains some of its capabilities through a scenario of telling travel stories. In this interface, the storyteller is able to:

1) Chat with a story viewer, a friend whom he would like to share the experience with, in plain text to "talk" about the stories (Fig. 1-2a bottom), see the matched media elements (Fig. 1-2a left), and edit the story by drag-and-drop of media elements to enhance his chatted story (Fig. 1-2a right),

2) Preview the photos and videos with captions (Fig. 1-2b), and

3) See Raconteur's suggestion panel (Fig. 1-2c), including the story patterns and the raw material of the photo, audio, and video repository.

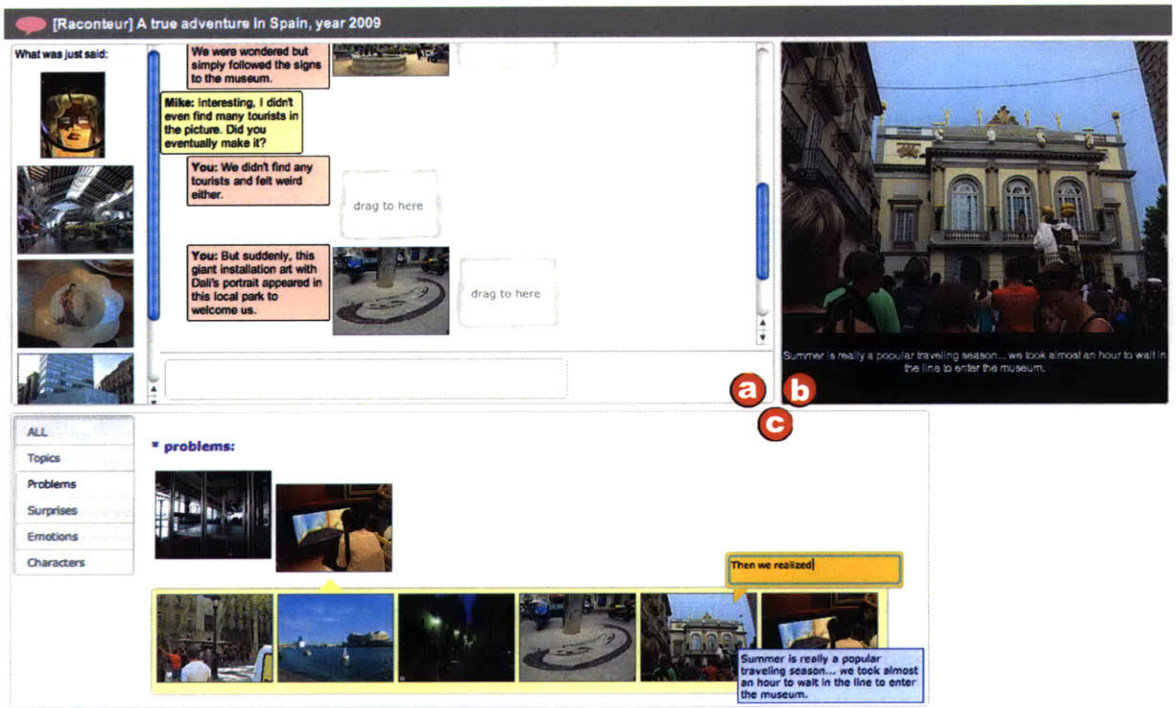

Fig. 1-2. Raconteur user interface, including: a) a chat box where storytellers can chat with a friend in text, see the matched media elements, and edit to enhance his story, b) a preview window to see the photos and videos with captions, and c) Raconteur's suggestion panel for observing the story patterns and the multimedia repository.

On the other side, the story viewer will see the same interface (Fig. 1-2a and Fig. 1-2b only) without the whole media repository, to motivate him to follow the teller and engage him to explore the unknown story. Meanwhile, Raconteur suggests questions for reference to the viewer. The goal of this interface design is to give novice users a sense of story creation and editing but empower them by putting in a familiar situation of chatting with a friend.

The final output of the system (i.e. the chat and edited result as Fig. 1-2a right) can be either a script with the selected scenes and users' narrations for later video editing, or a chat log for private use or sharing among friends.

Considering the following scenario:

(Beginning of the chat)

(Role)	**(Action and chat message)**
Teller:	[Input chat message in the chat box] *"My trip to Spain was full of surprising stories."* (Teller_message#1)
Raconteur:	[Suggest several story points with relevant media elements: finding installation art in a local park, visiting a police office, going to the tower of Gaudi's church, and seeing Asian products in a Spanish shop]
Teller:	[Select three topics from Raconteur's suggestions and drag the photos to Teller_message#1]
Raconteur:	[Update suggestions based on the edited files to show the potential story paths]

In our storytelling model, Raconteur asks the user (a storyteller) to compose his story with the user's friend (a story viewer) using dialogues. The storyteller can start the conversation by pointing out the overall story goal such as "*My trip to Spain was full of surprising stories*." The objective of the system is to provide a selection of possible matches to the story goal from the teller's personal corpus that best help to tell the story. Using Commonsense reasoning and analogical inferences, explained in detail later, Raconteur understands the concepts of "trip", "surprising", and "story", and reasons about the correspondence between the narrative goals and the concrete annotation. For example, the elements "finding installation art in a local park", "visiting a police office", and "seeing Asian products in a Spanish shop" are selected because their story sequences meet the goal "surprising stories". When the storyteller sees Raconteur's suggestions, he can select photos or videos of several topics he would like to share with his friend by attaching files to his chat message. Based on the edited elements, Raconteur tracks the story and updates the story paths.

(continued: viewer responds)

Viewer: [Click to chat about the photo of the installation art]

"Is that the art by Dali? Tell me more about the visit. I'm curious about how the Spanish culture that impacted his art work." (Viewer_message#1)

Raconteur: [Update suggestions of the precedent and following elements of this story point about Dali]

Teller: [Click to chat about the video taken in the train station]

"We wanted to visit Dali's museum, which was located in a city near Barcelona, so we needed to take a train there." (Teller_message#2)

Raconteur: [Suggest following elements: photos taken outside of the station, and photos of the installation art]

Teller: [Click to chat about the photo of a city view]

"The city looked peaceful but quiet, without anything of interest on the streets. We were wondered but simply followed the signs to the museum." (Teller_message#3)

[Drag one more photo of the city view to Teller_message#3]

Raconteur: [Suggest following elements: a photo of the installation art, a photo of waiting in a long line in front of the museum, and a video taken in the crowded lobby in the museum]

The viewer sees the teller's initial story and Raconteur's suggestions. He chooses one topic and responds by raising a question about the media element of "Dali". Raconteur matches his message to the story topic, and suggests the other media elements about this theme to the storyteller, to assist the teller in developing his point. The teller chooses to continue describing about his visit to the Dali Museum, but explained his first impression of the local city.

(continued: viewer responds and teller makes a point)

Viewer: [Click to chat about one of the photos of the city view]

"Interesting, I didn't even find many tourists in the picture. Did you eventually make it?" (Viewer_message#2)

Teller: [Click to chat about the photo of the art]

"We didn't find any tourists and felt weird either, but suddenly, this giant installation art with Dali's portrait appeared in this local park to welcome us." (Teller_message#4)

Viewer: [Input in the chat box]

"Wow! Now you can be sure you have come to the right place to see Dali's masterpiece!" (Viewer_message#3)

The viewer follows what the teller shares and finds the experience out of his expectation. Finally, Raconteur step by step helps the teller to make the point of "surprising stories" and create a story path that reflects both users' interests through this interactive process.

1.4 Design Challenges and Contribution

To achieve this goal of enabling conversational storytelling with personal captured media, this thesis confronts several design challenges concerning the conversation model, narration understanding, story pattern finding, and user interface design. It contributes the following aspects:

First, Raconteur creates a new interactive way to tell and edit personal stories with digital media by enabling and enhancing conversations between storytellers and the audience. This helps storytellers to brainstorm their life stories with a viewer that they want to share with, beyond the traditional story editing or composing environment that commonly allow only single users. Moreover, storytellers are able to create several kinds of story paths during the chats with different viewers that may reflect both teller and viewer's interests.

Second, in order to understand users' captions and narrations, and the stories behind them, it is challenging to integrate state-of-the-art technologies including natural language processing and commonsense reasoning with the study of conversational models. Our focus on story pattern analysis shows how an intelligent system can assist users in developing stories to make points. Beyond keyword search or topic spotting to enhance a single narration, our system considers story development and iteratively tracks the user conversation.

Third, our intelligent interface design focuses on enabling the human-human interaction through natural dialogue supported by our system, which serves as an assistant role instead of a conversational software agent that directly communicates with users and computationally tell stories. This helps storytellers concentrate on sharing their life experiences with friends, i.e. we put human intelligence at the center of the system to create conversational narrative as the final product.

Last but not least, Raconteur opens a design space that engages users in each other's stories based on intentional conversation. In addition to a one-time chat scenario, this design can be further applied to different scenarios such as multiple chats with various viewers, group story sharing among multiple users, unknown story exploration, or even the image capturing phase. After all, personal stories are to identify oneself and communicate with others. Maybe, this will motivate people to explore more about each other's life moments: to capture and share stories more often, more easily, and more enjoyably.

1.5 Thesis Organization

The rest of this thesis is organized as follows: in Chapter 2 we discuss the background of this research, including narrative theories, conversational storytelling, story analogy, commonsense reasoning techniques. Based on the introduced theories, Chapter 3 presents our two formative user studies. These lead to our system design with a detailed description of the structure, components, and implementation presented in Chapter 4, and the user interactions flow in Chapter 5. Then we move to evaluation and discussion (Chapter 6), followed by related work (Chapter 7). We end with our conclusion and future work in Chapter 8.

Chapter 2

Background

Storytelling is an essential part of everyday life. It may seem easy and natural for everyone to share their life experiences; however, telling *engaging* stories requires something more. To design an intelligent interface that assists people telling life stories, we surveyed relevant background from literary criticism and from artificial intelligence (AI). The four main areas we introduce are: narrative theory, conversational storytelling, story analysis, and commonsense computing. We explore how stories can be told, how important stories are, and how stories help people understand each other and reason about the world. This background knowledge supports our proposed new storytelling model and interaction design with digital media.

2.1 Narrative Theory

Long before the invention of writing, people have used various forms of narrative to share, communicate, and preserve the experience of daily life or in one's imagination. We continuously evolve the ways we tell stories, from prehistoric times to the current age of writing, from using traditional media to the advanced digital storytelling era. Storytelling has existed so naturally for such a long time that most people don't even notice it. Fortunately, scholars have been researching narrative using various approaches with a long history. The study of narrative theory, or *narratology*, is to understand the nature and structures of narratives. Researchers have identified the important basic elements that compose the narratives. In this section we introduce the definition of narrative and its components, interactive narrative, and the trend of creating narratives using personal digital media.

2.1.1 Narrative and Its Components

First of all, we define the concepts around stories. From narrative studies [Prince 2003] and [Abbott 2002], a "*narrative*" is the representation of a "*story*". It is composed by two essential parts: the content and the expression.

- The content plane, or more commonly known as a "*story*", contains an event or a series of events with existent(s):
    - "*Event*" presents a change of state that happens in a story, which can be an "action" (that occurs with the specific agency of an existent) or a "happening" (without the agency). It usually includes verbs and nouns that infer the changes and subjects, often with adjective or adverbs to attach more information. For examples, "We walked along the main street to the port." and "It started to rain all of a sudden." show an action and a happening respectively.
    - "*Existent*" or "entity" refers to an actor or actant who involves in the story events and may take actions. The more common term "character" refers to a human or humanoid entity. For examples, "I", "my friend", "Mike", or a bus driver.
- The expression plane, or so-called "*narrative discourse*" or "*discourse*", is the story as narrated, i.e. the form to present the content. The same content can be shown in different forms, such as reordering the sequence of events, using different tones, changing focalization or perspective, speed, etc. For example, the simplest way to present events is in chronological order by the timeline (the events happened earlier will be recounted first, followed by those happened later), while more complicated one can be as analepsis, a.k.a. flashback: after presenting the recounted story, introduce the events that happened at an earlier moment prior to the current time and space.

Therefore, the word "storytelling" usually refers to the activity of creating and presenting a story, i.e. from the content plane to the expression plane. To recount the story, a narrative needs a "narrator", who can be part of or out of the story world, to introduce the happening events, usually via "narration" (the verbal narrative). A "narratee", on the other hand, is the narrator's intended audience who listens to the story. With a careful design of the narrators and narratees, the actual author of the narrative can engage the actual audience in the story being told. Last, a narrative can be shown via different genres and media, such as printed text (novels, newspaper, magazines, etc.), verbal presentation (speeches, TV news, radio programs, interviews), oral conversation, drama, drawing, movie, etc.

2.1.2 Interactive Narrative

The invention of computers enables narrative to be presented in an even more diverse way. Traditionally, when the author is not creating stories interactively with the audience as oral storytelling, once he or she decides the discourse of a story, the narrative itself is not easily changed, i.e. the narrative is shown in a linear order along a certain path from the beginning to the end of the story, such as a book from the first page to the last one with an specific ending. On the contrary, using computer programs or hyperlinks over HTML pages in the digital world, a narrative can be changed in real-time to the audience. Furthermore, when its discourse is dynamically reconfigured, a computational narrative can be shown in a "non-linear" way depending on how the computer users or players navigate and explore a story by making choices among different story paths.

One of the examples is interactive fiction (IF), which is a software environment that takes textual input commands from users or players to explore the story world. The commands can be actions ("look", "talk to the guard", "get key") to control the story character to interact with other story elements like characters or objects ("book", "calendar") to see the information or possible actions. Such IF needs the author of the narrative to define the story world and the mappings, so that the software can parse the user input, match it to the defined world, and dynamically output the incremental story. In addition to text adventures, there are also video games or other media that provide visual feedback to enhance the storytelling experiences. For example, Façade [Mateas and Stern 2003] is a digital interactive drama that allows player to participate in a 3D graphical environment and interact with two virtual characters through conversations in English. Based on the user input, it intelligently models user's intention and generates different narrations and tension to continue the story, leading to various endings.

Interactive narrative shows an insightful way to engage readers by providing different paths and experiences. This design not only enables the audience to choose their preferences, but also helps them to anticipate and imagine the story following the choice point. We are particularly interested in the authoring systems of interactive narrative. Various research projects have provided insights of narrative creation: The storytelling and planning system "Universe" [Lebowitz 1985] models the story structure as a set of hierarchical plans and generates plot outlines based on the author's story goal. Riedl and León's [2009] story analogous generation system is able to analyze story structures and transform existing stories to a novel context. Cheong *et al.* [2008] presents an authoring interactive narrative framework to help users construct branching story structure. Montfort [2009] designed an interactive fiction authoring system "Curveship" for users to narrate and control the narrative world. The system distinguishes the design of "content" (the story) and "expression" (the discourse) in different levels, so that the narrator is free to describe events other than chronological order and change the focalization. Harrell [2006] takes an approach to interactive narrative based in cognitive science theories of imagination. By considering how concepts can be generated via the blending of other concepts, Harrell's GRIOT system interactively structures narrative events and incorporates generated content for narrative or poetic text-based or multimedia content [2009]. His design of architecture and event structure framework shows how a narrative can be decomposed computationally based on narrative and cognitive linguistic theories. Gervas [2009] presented a review of several interactive narrative systems and discussed models of computational creativity. Based on the comparison, he proposed several design issues of story creation, such as identifying the key elements about the creator, output, audience, etc. to be considered.

From the above systems, we found key elements in engaging the audience are: making story structures, setting up expectations, and encountering surprises that violate these expectations. This can result in an enhanced reading or viewing experience.

2.1.3 Moving Toward Digital Personal Narrative

The popularity of digital cameras, camcorders, and camera phones empowers the ways average users record their daily lives. Capturing stories no longer requires professional skills in complicated operating commands and programming. In addition to writing diaries, mails, blogs, or microblogs, it's more and more common to tell and share stories with the support of visual media such as digital photographs and videos, which more straightforwardly present the actual experience. However, as the new generation becomes accustomed to telling their personal life stories over online platforms and social networking sites such as YouTube and Facebook, the need for some assistance in organizing stories becomes critical.

A photograph is able to capture or present the moment of one event, while a raw video clip may contain one or more than one event along a clear, continuous timeline. Both often contain one or more existents, especially in the scenario of personal stories. However, the challenge lies on presenting relatively large numbers of events from a media repository in an intriguing way to attract the audience (the narratees) and to express the storyteller's (the narrator's, usually the same as the author's) model and perspectives.

Because of the required efforts of managing and editing a large set of material, only a small percentage of online users are willing to "create" their unique forms of narrative, while most users simply present stories by chronological order of events [Kirk *et al.* 2006]. To reduce the user effort, there is much research work on automatic organization or management of a multimedia system by considering the information in addition to image/video content itself. Cooper *et al.* [2003] designed a similarity-based analysis to cluster photos by timestamps and content. Joshi and Luo [2008] presented a method to infer events and activities from geographical information. Ames and Naaman [2007] investigated the motivations for people to annotate photos, and proposed a capture and annotation tool ZoneTag on mobile devices by providing geo-tags. Engström *et al.* [2010] studied media production systems that involve both live streaming media and recorded content, while the latter is annotated with footage and can be accessed and replayed in real-time for live scenarios.

However, most of the research work on automatic media organization focuses on analyzing the basic attributes such as time and location; few of them consider the overall story development and story-oriented thinking with digital media. We believe telling personal stories can be more interactive to help authors communicate with others.

2.2 Conversational Storytelling

We have surveyed the studies of life stories in daily conversations with structural and cultural analysis to explore the nature of conversational storytelling that is happening everyday in human life among the society. At first we interpret the concept of a "story" in a higher level from the social perspective. Polanyi [1989] defined stories as "specific past time narratives with a point" (p.20):

> Linguistic texts are produced to accomplish communicative aims. *Stories* are told to make a point, to transmit a message – often some sort of moral evaluation or implied critical judgment – about the world the teller shares with other people. Exactly what telling a story involves in this respect, can be gotten at somewhat indirectly by considering the *report*, often linguistically identical to the story in terms of events and state information, but different dramatically in impact. Any parent who has ever received a dreary *report* of the day's happenings instead of a *story* in response to a cheery "Well, dear, what happened in school today?" will testify to the difference.

That is to say, to make a story interesting enough to a listener, a storyteller needs to connect the events and communicate his/her own opinions. He or she should avoid presenting the stories without a remarkable points or reportable events that make it difficult to be remembered, retold, and therefore "dreary" [Labov 1997]. As we are considering the scenario of presenting personal life media, we found Linde [1993] had specifically defined what a "life story" is in a similar way as a coherent system (p.21):

> A life story consists of all the stories and associated discourse units, such as explanations and chronicles, and the connections between them, told by an individual during the course of his/her lifetime that satisfy the following two criteria:
>
> The stories and associated discourse units contained in the life story have as their primary evaluation a point about the speaker, not a general point about the way the world is.
>
> The stories and associated discourse units have extended reportability; that is, they are tellable and are told and retold over the course of a long period of time.

Conversational storytelling, or conversational narrative, is one of the common ways we express our life stories. It usually happens casually in our daily lives, and involves two or more participants, including at least one storyteller and one story listener (i.e. recipient). Moreover, Polanyi [1989] explained "turn-taking" between participants happens frequently and in an orderly manner because of the equality of daily conversation, unlike the speeches, lectures, or interviews. At the same time, Polanyi indicated the conversational storytellers "*are under a very strong constraint to make their utterances somehow coherent with what has been going on immediately preceding their talking*", and they must "recipient design" the stories, i.e. "*what is said must be tailored to the specific people who are the story recipients.*" On the other hand, a story listener "*must acknowledge that a story has been told by responding to it in some way which indicates acceptance of the fact that it was told and which demonstrates and understanding of what is was about.*" That is to say, conversational stories are created in a meaningful progression that involves both speaker and listener to maintain and continue the story topics. In addition, Schank *et al.* [1982] presented a theory of conversation comprehension to explain how a speaker's intent can be understood by a listener matching to possible "points".

People are accustomed to telling life stories in a face-to-face situation. In the digital world, similar forms of conversational storytelling also become increasingly popular through online chat, including instant messaging (IM) and conversations in a virtual world (e.g. Second Life or role-playing game environment) where users are able to input textual narrations to interact with other online clients. However, there is still a gap between "sharing" and "telling" the captured life stories with digital content that we need to fill in.

2.3 Story Analogy and Patterns

Furthermore, in order to assist users in continuing the stories in a conversation, we also surveyed the studies of story analogy, which is an important factor of structuring personal stories for a teller and reasoning about new stories for a listener. Story understanding requires connecting perceived story elements in a structured way. We are inspired by how humans understand an unknown story using analogies, which are partial similarities between different situations that support further inferences and can serve as a mental model to reason about a new domain [Gentner 1998]. In addition to psychology study, there are several research works that address the importance of narrative structure and analogy: From the sociolinguistics perspective, Labov and Waletzky [1967] analyzed structure of oral narrative of personal experience. Their overall structure includes: orientation, complication, evaluation, resolution, and coda. Some researchers have also addressed the concept of "story grammar" to support story composition by a set of rules [Black and Wilensky 1979] [Black and Bower 1980]. Schank [1991] proposed the idea of "story skeleton" to explain how we construct and comprehend a story under a certain structure to communicate with each other. Moreover, he suggested how the underlying story structure might alter the story listening experience (p.152):

> If we construct our own version of truth by reliance upon skeleton stories, two people can know exactly the same facts but construct a story that relays those facts in very different ways. Because they are using different story skeletons, their perspectives will vary.

In other words, storytelling is a process of conveying the storyteller's story model to the audience. Certain types of structures will help similar points be presented again and again. This process of developing a story helps the viewers to better understand the stories and to enhance their listening experience. For example, when a problem is addressed by the storyteller, the listener will naturally expect to know causality and resolution. Superficially different events may illustrate analogous themes, so the ability to make analogies helps tell a story in a coherent way.

2.4 Commonsense Reasoning

To recount stories efficiently in our daily conversations, we rely on a large amount of human knowledge called "common sense" to improve communication. This section surveys the literature of common sense and life stories, and presents the available computer technologies that support our system.

2.4.1 Storytelling and Common Sense

Common sense is a set of assumptions and beliefs that are shared among people in our everyday life. For examples, "An airport is used for travel", "Art is beautiful", and "You would smile because you are happy". Because it's based on what a group of people commonly thinks and agrees with, it has been long studied by social sciences; the sociologist Garfinkel [1967] defined "common sense" as:

> ... the socially sanctioned grounds of inference and action that people use in their everyday affairs and which they assume that others use in the same way. Socially-sanctioned-facts-of-life-in-the-society-that-any-bona-fide-member-of-the-society-knowsdepict such matters as the conduct of family life, market organization, distribution of honor, competence, responsibility, good will, income, motives among persons, frequency, causes of, and remedies for trouble, and the presence of good and evil purposes behind the apparent workings of things. Such socially sanctioned, facts of social life consist of descriptions from the point of view of the collectively member's interests in the management of his practical affairs. (...) we shall call such knowledge of socially organized environments of concerted actions "common sense knowledge of social structures."

This kind of knowledge is so obvious that we assume others know so that we don't need to explain explicitly. Based on this knowledge, we can understand the common life events and the properties of objects by observing and reasoning on our own. Garfinkel also described how common sense helps people interpret each other:

> ... for the everyday necessities of recognizing what a person is "talking about" given that he does not say exactly what he means, or in recognizing such common occurrences and objects as mailmen, friendly gestures, and promises.

Moreover, as we previously introduced the concept of coherent system for structuring life stories, Linde [1993] explained common sense is a special kind of such system that is transparent to most people and doesn't need to specially apply to. Consequently, to reason about the users' stories and further assist the storytelling process in our system, we consider incorporating a common sense reasoning method, as which Mueller [2006] defined:

> Commonsense reasoning is a process that involves taking information about certain aspects of a scenario in the world and making inferences about other aspects of the scenario based on our commonsense knowledge, or knowledge of how the world works. Commonsense reasoning is essential to intelligent behavior and thought. It allows us to fill in the blanks, and to predict what might happen next.

Researchers have been developing many different methods to reason about the world we know, including using logical and non-logical reasoning. We introduced the latter method based on a large knowledge base collected from web users.

2.4.2 OMCS Knowledge Base

To enable computers to understand our stories and "think" more like human in a similar way, we need to help computers acquire the common sense knowledge. From 1999, researchers have been collecting common sense knowledge from volunteers on the Internet to build a knowledge base called Open Mind Common Sense (OMCS). The knowledge is in the form of 20 or so kinds of two-place relations, as shown in Table 2-1. The online volunteers are asked to define different simple or compound concepts (as noun, verb, adjective, or prepositional phrases) using these relations. For examples, two concepts and their relation can be as: "*AtLocation*(art, museum)", "*PartOf*(sculpture, art)", "*HasProperty*(art, inspiring)", which means "Something you find at a museum is art.", "Sculpture is a kind of art.", and "Art is inspiring". Currently, the knowledge base in English has over a million assertions from over 15,000 contributors, while the knowledge bases in other languages including Chinese, Portuguese, Japanese, etc. are also expanding.

Relation type	Indication
IsA	What kind of thing is it?
HasA	What does it possess?
PartOf	What is it part of?
UsedFor	What do you use it for?
AtLocation	Where would you find it?
CapableOf	What can it do?
MadeOf	What is it made of?
CreatedBy	How do you bring it into existence?
HasSubevent	What do you do to accomplish it?
HasFirstSubevent	What do you do first to accomplish it?
HasLastSubevent	What do you do last to accomplish it?
HasPrerequisite	What do you need to do first?
MotivatedByGoal	Why would you do it?
Causes	What does it make happen?
Desires	What does it want?
CausesDesire	What does it make you want to do?
HasProperty	What properties does it have?
ReceivesAction	What can you do to it?
DefinedAs	How do you define it?
SymbolOf	What does it represent?
LocatedNear	What is it typically near?
ObstructedBy	What would prevent it from happening?
ConceptuallyRelatedTo	What is related to it in an unknown way?
InheritsFrom	(not stored, but used in some applications)

Table 2-1. The set of the defined relations that connect concepts in ConceptNet 4[3]

[3] http://csc.media.mit.edu/docs/conceptnet/conceptnet4.html#relations

This data collected from OMCS is then represented by ConceptNet in the form of a semantic network, and can be accessed and analyzed using computer programs [Liu and Singh 2004]. In this way, the connected concepts can be expanded. The project continues to evolve into the current versions of ConceptNet 3 [Havasi *et al.*, 2007] and ConceptNet 4, which improves acquisition of new knowledge and language structures.

2.4.3 AnalogySpace Inference Techniques

In addition to a large common sense knowledge base, we are also looking for the ability to reason about knowledge so that we can make sense of the textual information more efficiently and powerfully. AnalogySpace is a powerful tool for analogical reasoning [Speer *et al.* 2008] based on the OMCS project. AnalogySpace represents the entire space of OMCS's knowledge through a sparse matrix whose rows are ConceptNet concepts, and whose columns are features, one-argument predicates that can be applied to those concepts. A feature generally consists of one of the two-place relations together with another concept. Inference is performed by Principal Component Analysis on this matrix, using the linear algebra factorization method called "Singular Value Decomposition" (SVD). As Fig. 2-1 shows, by running a SVD on ConceptNet (as the original matrix A to be factorized), the space is transformed into a matrix of concepts and axes (the unitary matrix U), a diagonal matrix Σ of the axes, and a matrix of the features and axes (the conjugate matrix V). These axes are often semantically meaningful, and enable us to measure abstract concepts quantitatively by vector calculation, i.e. making the abstract concepts computable. For example, for two similar concepts (such as "dog" and "cat"), the value of the dot product of their vectors in row may be positive, indicating the two concepts share similar nature in many aspects (e.g. both dog and cat have 4 legs, are animals, can be pets, etc.). On the contrary, "dog" and "airplane" share different features (e.g. dog cannot fly as airplane, while the latter is not an animal, pet, etc.) therefore are not conceptually similar.

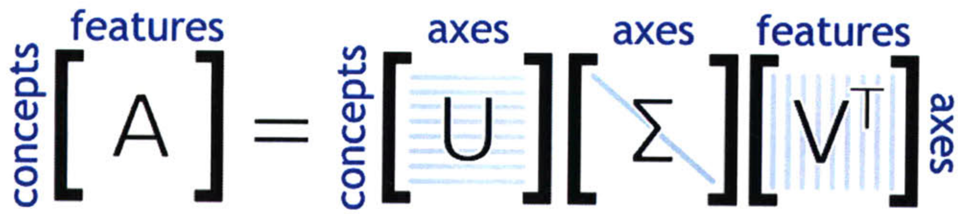

Fig. 2-1. Transforming a knowledge base into matrixes by SVD [Havasi 2009]

The reason this is good for computing analogy is that concepts that have similar Commonsense assertions true about them wind up close to each other in the transformed space. Unlike first-order logic approaches to analogy, it is computationally efficient, and tolerant of vagueness, noise, redundancy, and contradiction. Several important features that AnalogySpace provides for story reasoning include:

- Getting an ad-hoc category of a concept (e.g. "art", "museum", "sculpture" may fall into one category along with "painting" and "artist" as Table 2-2 shows),
- Measuring the similarity of different concepts (Are "art" and "park" conceptually related?), and
- Confirming if an assertion is true based on the current collected knowledge ("Are you likely to find art in a park?").

Concept	Similar concepts in an ad-hoc category
art	museum, sculpture, painting, artist, ...
park	city, grass, flowers, balls, animal, ...
travel	drive, transportation, go somewhere, take bus, fun, ...

Table 2-2. Examples of similar concepts found by AnalogySpace

In this way, we can provide users the freedom of describing their stories without word constrains. In addition, we can reason about the narrations and understand the inferred intentions, moving the system from word matching to story understanding, and most important of all, assisting storytelling.

2.5 Summary

In this chapter, we defined terms related to narrative and storytelling, discussed interactive narrative and its interaction design, and the need for assisting storytelling from personal digital media. Several background theories of life stories, conversational storytelling, and story analogy mainly from Linde [1993], Polanyi [1989], and Schank [1991], were introduced to support the chat design of Raconteur. Last, we introduced and discussed the important characteristics of a large commonsense knowledge base ConceptNet and the commonsense reasoning tool AnalogySpace, which will serve as the basis of Raconteur's knowledge and story understanding by identifying the relations of concepts in our living world.

Chapter 3

Formative User Studies

Prior to designing the Raconteur system, we conducted two formative user studies based on the background theories of narrative, conversational storytelling, story patterns, and commonsense reasoning. Our goal is to understand the user experience about: 1) conversational storytelling with the assistance in relevant personal digital media material and 2) media composition with the assistance in pattern analysis.

The system suggestions generated by the version of the prototypes used in this formative study are performed using simple text search only. The full Raconteur system presented later, and evaluated in Chapter 6, generates its suggestions with far more sophisticated natural language processing, Commonsense reasoning, and story pattern recognition. The simple text search is only used here as a baseline, so that we could gain experience with the effect of providing suggestions in the chat interface on the conversational process.

3.1 Study #1: Chat and Digital Media

In section 2.2, we introduced how people tell stories when conversing with another person. However, in the digital world using captured media files, can we also transform storytelling from a "*dreary slideshow*" to an "*engaging story*" by enabling a conversation? We conducted a small user study using a chat interface with a simple media search function to understand what two users would chat about over personal digital media.

3.1.1 The User Interface and Experimental Setup

For this study, we design an experimental user interface that contains a media repository and chat box, as Fig. 3-1 shows. Furthermore, we had built a basic string search function to match the narration with the captions of the media elements. In this interface, the storyteller is able to:

1) See the raw material of the photo, audio, and video repository (Fig. 3-1a) and preview the files,
2) Chat with a story viewer, a friend whom he would like to share the experience with, in plain text to "talk" about the stories (Fig. 3-1b), and
3) Compose the story by drag-and-drop of media elements (Fig. 3-1c) supported by the system's found files (Fig. 3-1d). The story viewer will see the same interface without the whole media repository (i.e. only Fig. 3-1b and c).

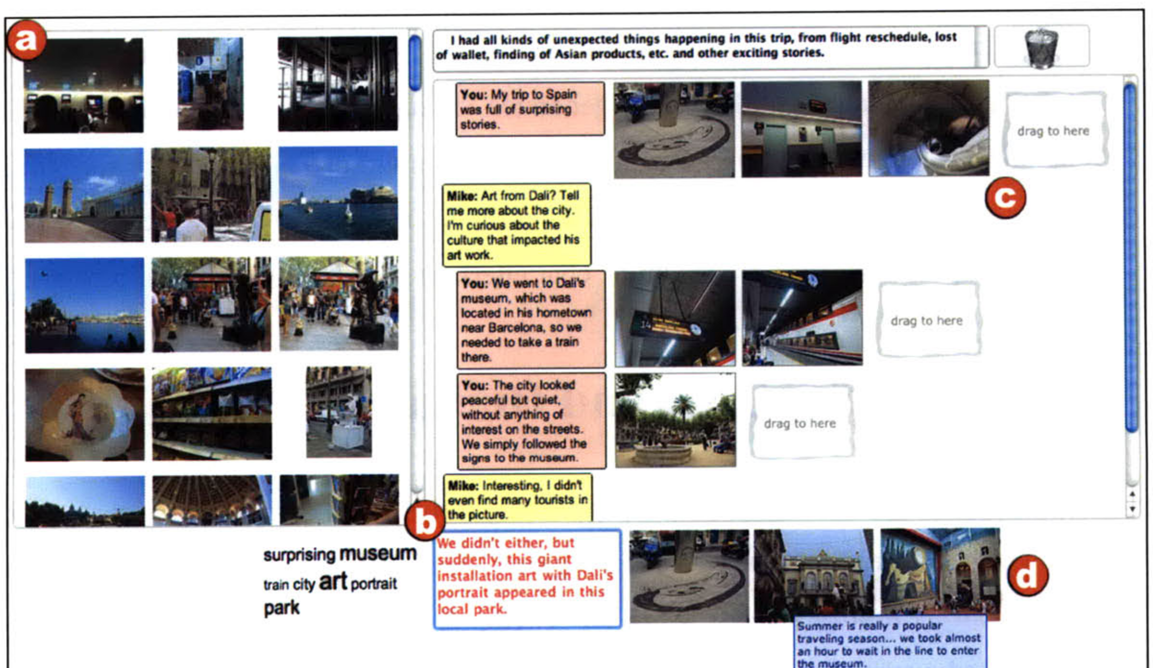

Fig. 3-1. User interface for the formative study of chat, where storytellers can: a) see the multimedia repository, b) chat with a story viewer, c) compose the media elements, and d) observe the basic search result.

We invited pairs of participants to chat about their stories. Each person took turns, first as a storyteller, then as a viewer. We asked participants to bring samples of their personal media files and orally tell the experiment facilitator their stories. The facilitator chose an appropriate set, according to the complexity of the collection. After the facilitator introduced our editing interface, each pair of the participants then would be asked to chat and edit the storyteller's material through the prototype.

We conducted this study with 6 participants, aged from 25-35 years old. The six story topics chosen were:

1) A five-day vacation in Spain,
2) A two-week business trip to Asia,
3) A one-week field trip to Italy,
4) A four-day visit to Boston,
5) A one-day biking trip in Cape Cod, and
6) A one-day conference organization and presentation in Singapore.

On average, the size of one repository was 69 media elements, including 95% still photos and 5% short video clips (most within 30 seconds).

3.1.2 Results and Discussion

The average story contained 60 messages. We observed the chat log and found: the conversational behavior of "turn-taking" happened during the online chat. The storytellers started the chat about their stories with the support of the system's suggestion, and then the story viewers read the chat messages, previewed the files, and posted questions or comments to interact with the tellers, such as "*what was the guy holding in the photo? I haven't seen such a thing in a conference before*" or "*You designed all that printed work? That's something very impressive!*" In other words, it was not the teller to tell stories alone as a "report", but an interactive process to illustrate a story together. All of the storytellers were able to respond to most of their friends' questions or comments by explaining with more story details. Moreover, there were several continuous story "points" being made by the tellers who presented their goals and background information through this process. For examples, "*Can you believe it? I finished my 10-mile biking trip! (...) You know, I don't exercise much...*" "*I went on this trip as one of the presenters. That's why we were busy setting up this giant poster wall*", and "*The light show was one of the famous spots in Hong Kong, so we were in a hurry to get on the ferry*" beyond simply describing the events captured in the photos or videos. In the post-test interviews, all the participants agreed that chatting through Raconteur was more "enjoyable" or "very enjoyable" rather than watching a slideshow from an online album. This small-scale study provides evidence for the value of conversational storytelling in the digital media editing process.

3.2 Study #2: Story Composition and Patterns

In this section we describe a formative study to understand how the display of story patterns found in a media repository may help users to edit, and to see if users would find presentation of analogous story elements helpful in story construction [Chi and Lieberman 2010]. The present study is concerned with the value of the analogical inference. We designed a story-editing interface that shows both the raw set of selected material and the analogous elements we found.

3.2.1 Story Collection

We asked one participant to collect media (photos or videos) documenting her life experiences for three months, and then asked one experienced facilitator to observe the 30 media collections and explain how she would structure the stories to compose into an integrated video. She selected and compared similar topics, and specified the key shots in each story. We summarize three main categories:

1) Collections with a clear procedure as a story pattern, e.g. birthday parties that people give surprises, sing the "Happy Birthday" song, make wishes, cut the cake, share gifts, etc.
2) Collections without a clear procedure but with certain expected events, e.g. graduation or farewell parties that people celebrate for a reason but the activities vary.

3) Collections without a clear procedure and without expected events, e.g. a camping, hiking, and biking trips that include difficult challenges and new experience of an activity. This especially applies to the travel scenarios.

We chose one media collection from each of these categories as test cases, including “Hsien’s birthday party with a potluck dinner”, “Mike’s commencement party for his first master's degree”, and “A 4-hour biking challenge in Cape Cod”. The collector annotated each media element with a sentence or two in English.

Then, we analyzed each collection to summarize the possible story patterns in it. Table 3-3 shows an example of the found pattern from the selected one-day biking trip: the element A1 with descriptions of “Cape Cod”, “stunning”, “famous”, “vacation”, and “biking” infers this piece of material indicates the user’s anticipation of the trip; P1 infers the preparation including having brunch, and renting a bike; D1 explains the difficulty of finding the way to avoid getting lost; then R1 shows the excitement of the arrival. We summarized all the patterns for users to navigate.

Anticipations and Worries
• "Cape Cod, a peninsula with stunning scenes, is famous for vacation and outdoor activities such as biking." (A1) • "Before this, I had only experienced long biking trip once. When I promised to take this challenge, I was a little bit nervous and afraid that I couldn't finish the trip." (A2)
Preparation
• "To start a day, we need energy; so first, we went to have a luxurious brunch." (P1) • "It is important to rent a good bike for the challenge. We came to the rent shop and pick our own bikes." (P1) • "Then, we biked all the way to the end of the rail trail. This was about six miles, the most difficult part." (P2)
Difficulties or problems
• "It took us a while to find the correct way between the branches. Thanks to the map, or we would get lost." (D1) • "It was a really long trail... I almost wanted to give up on the half way, especially I had no idea about how long I had biked." (D2)
Results (Resolutions): Successes or Failures
• "As you can see, we were really thrilled when we arrived the beach. Although I already felt one mile was long!" (R1) • "Surprisingly, it was easier than what I imagined. We were so excited when we arrived the destination." (R2)

Table 3-3. Multiple story units with similar patterns. The number indicates the elements in the same set of patterns.

3.2.2 The User Interface and Experimental Setup

To observe how such story patterns assist user in constructing stories, we designed a user interface for a single-user to observe story patterns and edit the media elements shown as Fig. 3-2, where the storyteller may:

1) See the unorganized, sequential material in chronological order (Fig. 3-2 a2).
2) Decide a story goal in English (Fig. 3-2 a1), and then the analogous elements will be shown (Fig. 3-2 b).
3) Drag and drop photos or video clips as desired to create a story (Fig. 3-2 a3).

Fig. 3-2. User interface for the formative study of story patterns: (a) the upper part presents the raw material of the unorganized collection (a2), and provides the editing interface for users to decide the story goal (a1) and the sequence of scenes (a3); (b) the lower part shows the sets of analogous story elements in a pattern that matches the story goal.

5 participants were invited, including 3 males and 2 females, aged between 20-30, experienced with digital media. Participants were asked to edit stories for sharing with their friends. The facilitator first helped them familiarize themselves with the test cases, and then introduced our editing interface and conducted the 3 editing sessions.

3.2.3 Results and Discussion

We found that when the size of the corpus was large and the contained story elements were relatively complex (test case 1&3), presenting the analogous story helped users follow a story pattern better. Especially for test case 3 (a biking trip), participants found the story complex, and reported the analogous examples helped them to design the story development. Most participants spent considerable time on observing the similarity of story content. One participant said, "*It was interesting to see how the system presented a new perspective to the story I wanted to tell*"; another explained, "*The system helped me rethink the similarity and differences between experiences, which I would rarely think to do from just browsing a bunch of files.*" These findings encouraged us that the analogical reasoning mechanism would prove useful to users in story construction.

Our formative user study also shows that this kind of analogy finding is particularly helpful in the case where users have large libraries or complex material, especially for travel scenarios. This presentation encourages users to think about the story goal instead of directly composing individual elements. Moreover, this design might be helpful with creative discovery to present stories from different perspectives, in addition to simplifying the story construction process. Therefore, to assist users presenting their stories, we not only need to find out the media elements with similar topics, but also the underlying story structures.

3.3 Summary

We conducted two formative user studies to understand the nature of the chat behavior with digital media and the potential for assistance with story patterns. We confirmed that in the digital world, the key features of conversational storytelling can still be observed in online chats, such as turn taking, making story points, and questioning and answering, similar to our daily face-to-face conversations. These support our design to enable a chat between a storyteller and audience to create and edit life stories together, and to provide structural story patterns for users to consider.

Chapter 4

System Design and Implementation

Based on the above background study and observations, in this chapter, we introduce the system design and each component of Raconteur.

4.1 Raconteur Structure

We designed the system to reason about stories from a personal multimedia repository for users to interactively chat and edit. Fig. 4-1 shows Raconteur's system structure, which is composed of several major components as follows:

- A multimedia database of multiple media elements that are annotated with textual information,
- A narration processor that parses the user's narrations and captions,
- An analogical inference model and a story developer that connects to a common-sense knowledge base, and
- A user interface that allows a pair of users (a storyteller who owns the multimedia data and his/her friend as a story viewer) to chat about the story, observe the system suggestions, and edit in real-time.

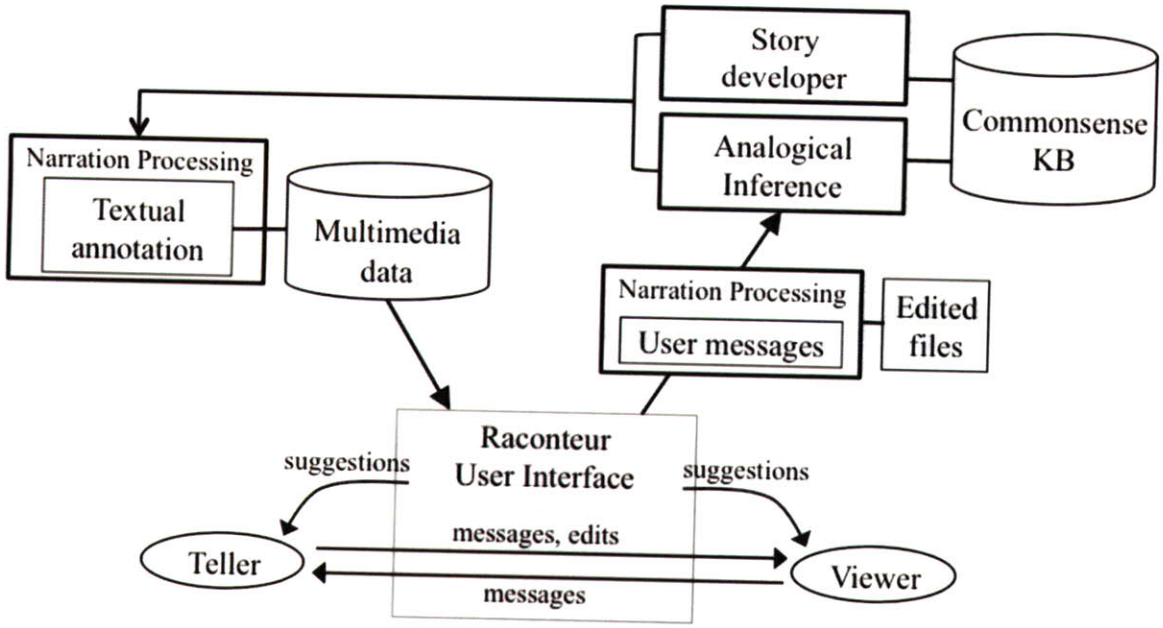

Fig. 4-1. Raconteur system structure

To assist the users' conversation about the captured multimedia material, Raconteur tracks the chat messages and the edited files, and updates suggestions to the user interface in real-time by computing the user narrations to match with the annotated media elements. We compute the following information for each user narration and each element annotation:

1) The narrated concepts from sentence structure,
2) Their concept vectors in AnalogySpace,
3) The story elements including characters and locations, and
4) The relations with other elements, i.e. the story patterns.

The following sections introduce how the stories can be reasoned based on this information.

4.2 Resources of Multimedia and Assumptions

For a given multimedia repository, we see each photo, video clip, audio file, or other media, all as an individual "media element", i.e. story unit in the system. We assume that each of these elements is annotated with a sentence or two in unrestricted natural language. The annotation may describe characters, events, and intents of the captured scene. For example, as Table 4-1 shows, *"This installation art by Dali showed up on the way to the museum. It was a big surprise because we didn't expect to see this in such a local park."* We are looking for the information in such a higher level other than simply subjects, objects, emotions, etc. The objective of this is to acquire the intent behind these media elements, the events happened, and the contextual relationships between existents. We could also obtain some of information from other sources like tagging (individual words), location data, face recognition, or object recognition by image processing. In most cases, annotations are explicitly provided by users, while some annotations may be generated by metadata, transcription of audio, or other means.

Such a repository can come from a personal content management system that enables users to attach textual annotation to files, or any online media collection platform accessible through Application Programming Interfaces (API) such as Picasa, Flickr, Facebook, or YouTube that allow users to maintain personal multimedia data and edit information including captions or media summary. The Raconteur system needs to access users' album list, titles, dates, descriptions, and the contained lists of files (photos and/or videos), each with file system links or hyperlinks of thumbnails and content in different sizes, and the information of captions, file types, date, etc.

In our design, any of the unannotated multimedia elements will be kept in the repository but not considered by the analysis. However, they can be referred to and attached if users so specify explicitly. The narration during the chat on these elements may also be considered as additional information for future references of different chats to enrich the understanding to the repository.

type	**Photo**
shot	
caption	"This installation art by Dali showed up on the way to the museum. It was a big surprise because we didn't expect to see this in such a local park."
type	**Video clip**
selected shots	  
duration	1'00" (1 minute)
caption	"Two singers were performing the famous aria "None Shall Sleep" from the opera "Turandot" in this street corner in Barcelona. Again, art can be so close to daily life."
narration	(The author talks to the friend:) "That man just walked from the audience to sing with him? Amazing!" (Music and singing) … (Audience applauding and cheering)

Table 4-1. Examples of stories behind a still photo and a short video clip taken from a trip.

4.3 Narration Processing and Representation

Raconteur analyzes both the annotation of each media element in natural language and the users' chat messages in real-time. This requires the natural language processing module and additional mechanisms that consider the semantic meaning in the story world. Our goal is to break the user narrations down to propositions and clauses by parsing the sentence structures, and then remove those non-story-world clauses so that we can focus on concepts that describe the stories for later story analysis.

4.3.1 Natural Language Processing

To understand users' input narration during the chat and the annotations placed on individual scenes, we parse all the textual information using the state-of-the-art natural language processing (NLP) tools. NLP is a research area of computer science and linguistics that applies computational methods to analyze human natural languages. It understands the grammar, the sentence structures, the possible intentions and the basic forms of words, and other tasks. In our design, NLP helps Raconteur to identify important concepts related to the stories. We applied the Natural Language Toolkit (NLTK) [Bird *et al.* 2008], a suite of programming libraries for symbolic and statistical NLP. It's capable of analyzing English and several other languages. We particularly use several features:

5) Part of speech (POS) tagging to identify words including verbs, nouns, and adjectives/adverbs listed as Table 4-2, which may contain possible contextual information to illustrate the stories. In addition, we also consider conjunction markers in conversation to identify the intention of sub-phrases, such as "because", "however", "in order to", "anyway", etc., which may indicate reasons, transitions, purposes, and other connectives.
6) Named entity recognition (NER) to determine story characters (names like "Peter", "Gaudi", "Dali"), organizations (e.g. schools, museums), geographical areas (e.g. "Spain", "Barcelona"), and time (e.g. "one hour", "July 4th") that help categorize the basic story elements.
7) Stemming and lemmatization to normalize words into the basic forms (e.g. "went" into "go", "the cars" into "car"), for the later concept processing and comparison.

Tag	Meaning	Examples
ADJ	adjective	new, good, high, special, big, local
ADV	adverb	really, already, still, early, now
CNJ	conjunction	and, or, but, if, while, although
DET	determiner	the, a, some, most, every, no
MOD	modal verb	will, can, would, may, must, should
N	noun	year, home, costs, time, education
NP	proper noun	Alison, Africa, April, Washington
NUM	number	twenty-four, fourth, 1991, 14:24
PRO	pronoun	he, their, her, its, my, I, us
P	preposition	on, of, at, with, by, into, under
UH	interjection	ah, bang, ha, whee, hmpf, oops
V	verb	is, has, get, do, make, see, run
WH	wh determiner	who, which, when, what, where, how

Table 4-2. Selected simplified POS tagset in NLTK[4]

Table 4-3 shows an example of the narrative sentence in the previous subsection being processed by NLTK to identify the verbs ("show"), nouns ("installation art", "way", "museum"), and name of a person ("Dali"). In this way, all the narrative sentences will be decomposed into a structure of potential concepts and phrases for later analysis. However, please note that negative sentences such as "without", "nobody", "never", "nothing" that will alter the semantic meaning are not currently identified by the system, but will be considered as future work.

[4] http://nltk.googlecode.com/svn/trunk/doc/book/ch05.html

Original sentence	Processed sentence with POS-tags and NER	Processed sentence into basic forms
"This installation art by Dali showed up on the way to the museum."	(S This/DT installation/NN art/NN by/IN (PERSON Dali/NNP) **showed/VBD** up/RP on/IN the/DT way/NN to/TO the/DT museum/NN ./.)	(S This/DT installation/NN art/NN by/IN (PERSON Dali/NNP) **show/V** up/RP on/IN the/DT way/NN to/TO the/DT museum/NN ./.)

Table 4-3. The processed sentence structure as a result of natural language processing

4.3.2 Non-story-world Clause Removal

From NLP, we can identify the interjections or reinitiation markers that are not referred to things happening in the story world but often used in conversations, such as "yeah", "Gosh", "oh", etc. However, we also need to remove the non-story-world clauses that contain verbs but could not provide story-related information, such as "think", "mean", "know", "guess", etc. Polanyi [1989] explained (p.21):

> Stories are highly complex discourses, however, and not all the propositions about the storyworld are equally important to the point which the story is being told to illustrate. (...) only the events which cause alterions in those states bring about a contrasting state which is "meaningful."

Table 4-4 lists more detailed examples of such clauses from Polanyi's research studies of American oral conversations. By doing so, we can reduce the concepts that are too general that may associate with too many other concepts. Note that we only remove such short clauses, but keep those following sub-clauses, and analyze all the potential story world propositions that may contain story information and opinions for story understanding.

Type	Examples
Interjection	Yeah, god, gosh, oh, huh, uh, man, well, so, right, yes, ...
Non-story-world clause	I think, I mean, I said, I guess, I did, you know, you mean, you see, You wouldn't believe it, that's all, ...

Table 4-4. Lists of different types of non-story-world words and clauses in conversations

4.4 Analogical Inference

After the narration is being processed and its sentence structure determined, we then analyze relations between the events behind various media elements. We first build concept vectors for each element, look for analogies with various patterns in a repository, and then find the possible story sequences.

4.4.1 Building Concept Vectors

First of all, in order to measure the semantic meaning of narrative sentences and find patterns, we apply the common sense computational method of AnalogySpace inference, introduced in Chapter 2. Based on the result of NLP, we traverse each word of verbs, nouns, adjectives, and adverbs as a potential concept that may indicate events and story elements, such as "show", "art", and "inspiring". We look for the information by accessing the "vector" that computationally represents such a concept from the unitary matrix U with concept and axes in AnalogySpace. By doing so, we transform abstract semantic concepts contained in each element into a list of vectors that are computable for later analysis. For example, the narration in Table 4-3, "*This installation art by Dali showed up on the way to the museum*" that contains concepts of ("installation", "art", "show", "way", "museum"), will be represented by vectors of ($v_{\text{installation}}$, v_{art}, v_{show}, v_{way}, v_{museum}).

4.4.2 Associating Similar Elements

An important aspect of the system is to associate media elements that address similar story points to help users reason about a large set of material in a repository. Therefore, we measure the similarity by the concept vector calculation containing in the story elements.

The simplest measurement is to compare all the concepts of the annotations placed on two elements. Similar to the previous example of comparing two concepts "dog" and "cat" by their concept vectors in Chapter 2.4.3, we can compare two element annotations using their "narration vectors", as Fig. 4-2 shows: For each element represented by a list of concept vectors $V = (v_1, v_2, \cdots, v_M)$ captured from the annotated narration sentences, we add up its vectors into a single computable vector $V' = \sum_{i=1}^{M} v_i$. Then, we normalize this summed vector $\hat{V}' = \frac{V'}{|V'|}$ in order to scale the vector by its length so that we can provide the same basis for narrations of different lengths and different numbers of concepts. In this way, we can compare two elements by getting the "dot product" of their normalized vectors $s = \hat{V}_1' \bullet \hat{V}_2'$ to measure the similarity by narrated concepts. We examine the final value of the dot product to compute the similarity between the sentences: if the value is positive, the two elements are conceptually similar. This computation enables us to classify all the media elements to connect different events and sort by relevance. For examples, elements that contain concepts of "art", "museum", "gallery", "sculpture", and "inspiring" will be classified in a art-related category, while elements about "be stolen", "thief", "anxious", "police office", "report", will be categorized as another theft-related one.

Given a element with narration sentences $S_1 = (c_1, c_2, \cdots, c_M)$ with M concepts and another element with sentences $S_2 = (c_1, c_2, \cdots, c_N)$ with N concepts:

1. Transform S_1 and S_2 respectively into a list of concept vectors $V_1 = (v_1, v_2, \cdots, v_M)$ and $V_2 = (v_1, v_2, \cdots, v_N)$, where v is the concept vector of concept c from the unitary matrix U of AnalogySpace.
2. Add up the concept vectors respectively: $V_1' = \sum_{i=1}^{M} v_i$ and $V_2' = \sum_{j=1}^{N} v_j$
3. Normalize each vector: $\hat{V}_1' = \frac{V_1'}{|V_1'|}$ and $\hat{V}_2' = \frac{V_2'}{|V_2'|}$
4. Take the dot product of two vectors: $s = \hat{V}_1' \bullet \hat{V}_2'$
5. If $s > T$ where T is the threshold greater than 0, S_1 and S_2 are similar. Otherwise if $s \le T$, S_1 and S_2 are not similar or the relation cannot be determined.

Fig. 4-2. Algorithm of matching two media elements in the simplest way

Using concept associations, we can also generalize the user's statements so that users do not need to describe the events precisely or with structural constrains. Again, note that this is different from keyword expansion such as WordNet [Fellbaum 1998] that finds synonyms and synsets with lexical relationships (e.g. "buy" and "purchase", or "beautiful", "pretty", and "lovely" are lexically similar). Instead, it's possible to use commonsense reasoning to identify conceptual relations that may involve causality [Kuipers 1984] and other connections, such as "buy" and "wallet", or "beautiful" and "painting." ConceptNet that particularly includes the relations of "Causes", "CausesDesire", "HasSubevent", "HasPrerequisite", and "UsedFor" helps AnalogySpace to perform such inferences.

However, considering story events and elements, for some circumstances we do not simply calculate all the concept vectors, but identify certain key concepts that may be particularly related to story topics (especially verbs and nouns) or the concepts that infer further intention (e.g. "friends" implies story characters; "wallet got stolen" implies a problem). We will show more examples in the next section.

4.4.3 Story Pattern Reasoning by Making Analogies

To reason about larger patterns between scenes for users to structurally develop the story instead of chatting promptly without connections, we develop an analogical inference technique considering several patterns. In Chapter 2.3, we introduced the concept of story patterns, which are the structure that makes similar points. Telling stories by making such enhanced points usually helps story listeners to understand and follow the storyteller better, and each story path may provide different story experiences to the audience. Therefore, our goal is to find the elements with *connected events and similar intentions*.

4.4.3.1 Patterns by Problem and Resolution

Based on the collected stories from our formative studies (Chapter 3.2 and Chapter 3.2), we found the most common story pattern is encountering unexpected problems, especially in the travel scenario. This often makes a personal story "special" and impressive to the audience because it arouses the listeners' curiosity or reminds their similar life experiences. Table 4-5 shows some examples of such pattern from different stories we collected and analyzed based on the user annotation using our defined terms of *intention, problem, resolution, and consequence*. Examples include: the story "one-week trip to Spain" contains "buy living goods in a local market" (intention), "wallet got stolen" (problem), "report to the police" (resolution), and "cannot enjoy buying souvenirs" (consequence); the story "the first camping trip" contains "put up the tent" (intention), "trouble with composing elements" (problem), "reading instructions" (resolution), and "successfully settling down together" (consequence).

Intention	*Problem*	*Resolution*	*Consequence*
take a flight to Spain	flight was delayed	take the next flight	arrive the hotel late and feel exhausted
move to Cape Cod	get stuck in a traffic jam	-	arrive the destination but late
take a bus to downtown	unsure about the destination	-	arrive the city
buy a ticket to the museum	wait in a long line	-	observe the art work in the museum
buy living goods in a local market	wallet got stolen	search for wallet, report to police	cannot enjoy buying souvenirs
head to the destination	get lost	find a map	arrive the destination
bike	want to give up on the long trail	take a rest and continue biking	arrive the destination
enjoy the beach	start to rain	leave the place and go back	return to the rent shop
walk in the city	feel very hot	eat ice cream	-
take a dinner at a restaurant	unreadable menu	look for guide book, order food	enjoy dinner
put up the tent	trouble with composing elements	read instructions	successfully settle down together

Table 4-5. Lists of matched examples to the pattern of problem and resolution

From the table we can observe that the common feature of these "problems" include those concepts that people don't like, such as "delay", "traffic jam", "wait", "steal", "lose", etc. To detect this kind of concept, we reason using AnalogySpace: from the conjugate matrix V of features and axes, we acquire the vector $v_{person-desire}$ by querying the row vector of "*Desires*" with the concept 'person' on the left, which means the known concepts related to what a person desires or does not desire. Then, we compare the concept vectors from annotations with this desire vector by their dot product, so that a negative value indicates an "undesired" concept, compared to other positive concepts that people prefer such as "travel", "famous", "relax", etc. (Table 4-6). This inference enables us to identify those possible problems in a repository.

Problem related concepts	dot product value	Non-problem related concepts	dot product value
delay	-0.992	travel	0.018
traffic jam	-0.993	famous	0.687
wait	-0.243	relax	0.022
steal	-0.032	earn	0.025
lose	-0.110	win	0.017
rain	-0.457	sunshine	0.695

Table 4-6. Dot product results of the desire vector and exemplar vectors of concepts, where the negative value indicates people do not like that concept as much as other concepts with positive values

Given a media repository $R = \{E_1, E_2, E_3, \cdots, E_M\}$ such that all the narration sentences of each of the *M* media elements have been transformed into concept vectors $V_i = (v_1, v_2, \cdots, v_N)$, where *N* is the number of concepts in the individual element E_i:

1. Build a vector $v_{person-desire}$ from the conjugate matrix *V* of the AnalogySpace.
2. For each media element E_i with concept vectors $V_i = (v_1, v_2, \cdots, v_N)$, take the dot product of two vectors: $s_j = v_{person-desire} \bullet v_j$. If $s_j < T$ where T is the threshold less than 0, add E_i to the problem set P.
3. For each media element E_k in the problem set P, find the other elements E_l that associate with the concepts of E_k from the repository R and add each E_l into the group set G_k.
4. For each of the associated elements E_l in the group G_k, determine the relations (causality, subject or topic related) with the problem element E_k. Remove E_l if it's not semantically related. Add E_k to G_k.
5. Output each G_k as a member of possible collection of this pattern.

Fig. 4-3. Algorithm of finding a collection of problems and resolutions

After identifying the potential problems happening in the stories, we then reason about the connected events related to each problem, including the intentions, the resolutions, and the consequences. These events can include causality relations, or simply around the same topics or with the same subjects. Fig. 4-3 shows the simplified algorithm of this pattern matching, and Table 4-7 shows some examples of Raconteur's results. In this way, Raconteur finds the relations between the media elements, so that when the storyteller chats about any of the issues or individual elements, the system can provide suggestions to assist him thinking about the story development.

Intention	***Problem***	***Resolution***	***Consequences***
			
"After arriving the **destination**, we decided to **move** to another **beach** nearby." (image)	"We saw the **cloud** coming from the other side of the sea. We were **worried** if the weather would get **worse**." (video, length 00'13")	"We could only **leave** the **beach** and headed **back** soon." (video, length 00'05")	"It was not easy to **bike** **back**. Eventually we did it before the rental shop closed. This is the final picture with our rented **bicycles**." (image)
			
"On the first day in Barcelona, we saw a local **market** next to our hotel and decided to **buy** some living **goods**." (image)	"This is where I **lost** my wallet: sadly, I should have noticed the **thief** could **steal** my things in such a local market without effort" (image)	"You know where it is? We were at the **police** **office**... first time to visit such a place in my life. We need to report the **lost**." (image)	"We went to a **food** **market** where all the **fruit** looked fresh and delicious, but I better not to **buy** too much..." (image)

Table 4-7. Selected Raconteur's results from different repositories to the pattern of problem and resolution

4.4.3.2 Patterns by Expectation Violation

We have also found a similar pattern that produces the experience of surprise by presenting a violation of expectations or observations. Identifying the expectation violation pattern requires looking at several ConceptNet relations, not just a single relation like "Desires". Table 4-8 presents some of such examples: for two elements containing the same concept related to "park", one said "*On the way to the museum, we walked through a local park*", and the other describes, "*The installation art suddenly appeared in this park.*" We pose a question to AnalogySpace: "Is it likely to find *art* in a *park*?" If the result is negative but the two elements illustrate the same topic, we regard it as a match to this pattern. Establishing the expectation and showing violations helps users address the special moments they encountered and make memorable story points to the viewers [Schank 1986]. It also helps users structure a narrative to present events with connected, causal relation.

In addition to commonsense reasoning, from the grammar structure, we can also identify this kind of connection if the user explicitly describes it according to an assumption grammar, such as "*We thought there must have been full of tourists on the beach, but it was surprisingly calm with only a few families when we reached there.*"

Expectation (or observation)	***Violation***
walk through a park	installation art appears
see peaces of furniture	see the furniture composed into a face from another angle
go shopping for souvenirs	find famous products from foreign countries
walk on the street in Spain	see a Japanese restaurant
take a metro	find a couple with roses
full of tourists on the beach	only few families

Table 4-8. Lists of matched examples to the pattern of expectation violation

4.4.3.3 Patterns by Topics

We observe that continuing a story with connected topics helps an audience explore the story according to a certain perspective. For example, when talking about a conference or meeting, similar ideas such as organizers, presentations, posters, audience, etc. are often addressed. For another example, a trip to a city famous for art may include several stories like visiting an art museum, interacting with street art performance, going to a concert, etc. Therefore, we categorize all the elements in a repository by associating the elements with each other using the algorithm in Fig. 4-2.

4.4.3.4 Patterns by Emotions

Considering emotion is one of the important factors that alter the story experience. We identify several common types of emotion to analyze the repository. As Fig. 4-4 shows, we create vectors of "happy", "relax", "excite", and "worry" from AnalogySpace and match with media elements. Table 4-9 presents some of such examples.

Given a media repository $R = \{E_1, E_2, E_3, \cdots, E_M\}$ such that all the narration sentences of each of the M media elements have been transformed into concept vectors $V_i = (v_1, v_2, \cdots, v_N)$, where N is the number of concepts in the individual element E_i:

1. Build emotion vectors $(v_{happy}, v_{relax}, v_{excite}, v_{worry})$ from the unitary matrix U of AnalogySpace.
2. For each media element E_i with concept vectors $V_i = (v_1, v_2, \cdots, v_N)$, take the dot product of each emotion vector and the summed concept vectors: $s_k = v_{emotion-k} \bullet \sum_{i=1}^{N} v_i$. If $s_k > T$ where T is the threshold greater than 0, add E_i to the emotion set $Emotion_k$.

Fig. 4-4. Algorithm of finding collections bringing different emotions

emotion	***examples***	
Happy	 "We were extremely **lucky** to enter the room again before the museum closed to see the **famous** art work!" (image)	 "It was so **sweet** that the organizer offered each of us a cup of coffee for **free** to **wake up**!" (image)
Relaxed	 "What a **blue** and beautiful **skyline** and clean **beach**. " (image)	 "All these pillars to support the church are designed as **trees**. Look like they are **alive**." (image)
Excited	 "We went to a Sepak Takraw game. It's a **sport** like **volleyball**, except you can only **hit** the **ball** with your feet or head, like **soccer**." (image)	 "There are always **surprises** in a **new journey** - look at what we found, a turtle crossing the street!" (image)
Worried	 "This was the time of the swine flu **scare**, so many people wore masks. But nobody got **sick**." (image)	 "Steven was looking at the wall with an **unsatisfying** look on his face. 'It's too **curve** for a wall, isn't it?' He said." (image)

Table 4-9. Lists of examples to the pattern of emotions

4.4.3.5 Patterns by Characters and Locations

Considering what Polanyi [1989] explained about how storytelling works in a conversation (p.15), it is important to help the viewer to understand the background information and basic story elements of the story world when the speaker tells a story happened in different context:

> ... story recipient are alerted by conventional story introducers which a would-be storyteller uses to signal the intention to tell a story. The talk then moves out of the here and now of the conversation into a storyworld: another time, often another location, populated by other participants.

Introduced in Chapter 4.3.1, Raconteur is also able to identify characters and locations by named entity recognition. Found items include human names such as Mike, Tom, and Ann, and geographic names such as Cape Cod, Japan, and Hong Kong. However, to assist users in developing stories, we track not only known names, but also the abstract concepts around characters and locations in the narrations. For example, when a user says, "*I went on this trip with several of my friends*," using AnalogySpace we understand the word "friend" refers to "people", and particularly select those media elements annotated with characters' names or similar concepts, such as "*Jacky and Mike were asking for the direction*" and "*Our group photo with the famous landmark*".

4.4.4 Finding Story Paths

Finally, for each collection of elements from the found patterns, Raconteur reasons about one or more compelling story sequences considering causality and time factors. We decide whether two scenes are connected because of causality, e.g. each pair of "get – ticket" and "enter – gallery", "enter – gallery" and "see – portrait", "lose – thing" and "go – police office", "go – restaurant" and "read – menu" can be sequential. Otherwise, we sort the elements linearly along the time line.

4.5 Story Developer

After the repository is analyzed, Raconteur keeps track of the overall story development and suggests media elements using a planner to help users present the main point of the story in real-time. Our story developer maps the user narration to the pre-analyzed story patterns and updates the connected events as causal paths. It detects the user edits that match to the paths, and avoids frequent suggestion of the same elements that have been edited and shown.

4.6 Implementation

Finally, users can tell stories and interact with friends through Raconteur's user interface. Fig. 1-2 in Chapter 1 shows the overview of our web user interface, which can be accessed by any common web browsers without installing additional plug-ins. It is implemented using HTML (HyperText Markup Language) with CSS (Cascading Style Sheets), the JavaScript language, the jQuery[5] JavaScript library, and XML (Extensible Markup Language). The web user interface sends requests to Raconteur's web server, which connects to the main program (programmed in Python) that analyzes the user input and then outputs the matched results of media elements and suggestions back to both the users' web interfaces to update in real-time. The Raconteur main program applies several toolkits including NLTK[6] for narration processing, ConceptNet[7] and Divisi[8] (which contains AnalogySpace) for commonsense reasoning, and Picasa Web Albums Data API[9] powered by Google for accessing user's online media repository. In the next chapter, we introduce the components of the user interface and the user interaction in detail.

---

[5] http://jquery.com/

[6] http://www.nltk.org/

[7] http://csc.media.mit.edu/conceptnet/

[8] http://csc.media.mit.edu/divisi/

[9] http://code.google.com/apis/picasaweb/

Chapter 5
Chatting through Raconteur

This chapter introduces the Raconteur interface and its components, including a chat box, preview window, and suggestion panel, as Fig. 1-2 on page 25 shows. We introduce several aspects of the user interaction: how a user prepares for a conversation with personal media repositories, chats about the stories with a friend, and observes and edits with Raconteur's suggestions.

5.1 Preparing for Chat

To chat about stories through Raconteur, the storyteller uploads his or her captured personal media to an online photo/video-sharing website such as Picasa and annotates the files by adding captions through its web interface (Fig. 5-1). We assume each online album contains only material from a single episode or activity, for example, a one-week trip.

When the user logs in to our system given a Picasa username, Raconteur will allow the user to choose an album for the chat and invite a friend (Fig. 5-2).

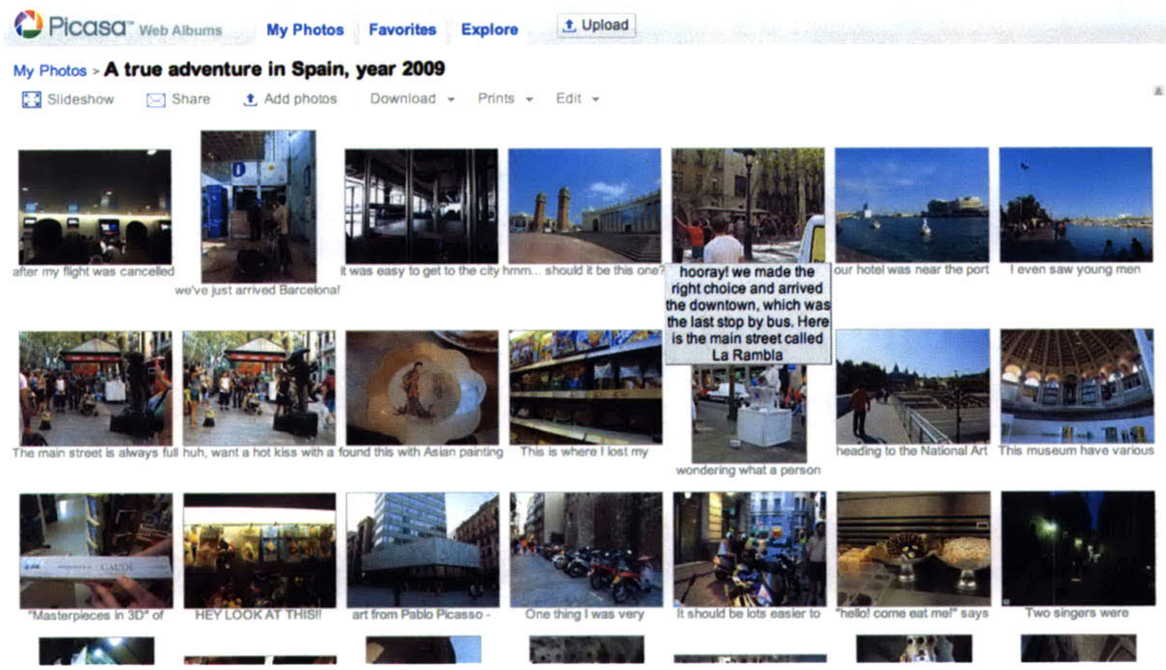

Fig. 5-1. User uploads media elements to an online album and annotates with captions.

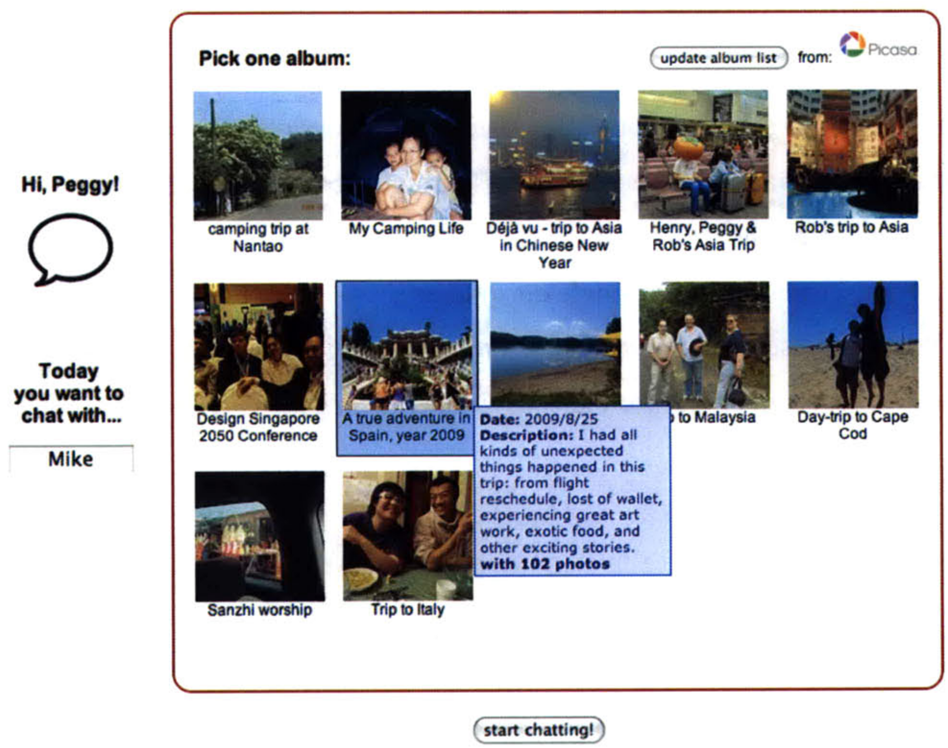

Fig. 5-2. Raconteur's login interface: by reading the online album information, the system allows a user to choose a story to chat.

After logging in to Raconteur with a chosen set of material, the user will see the initial interface as Fig. 5-3. He can browse all the elements, preview any image or video, and see the annotated captions using mouse hover (Fig. 5-4).

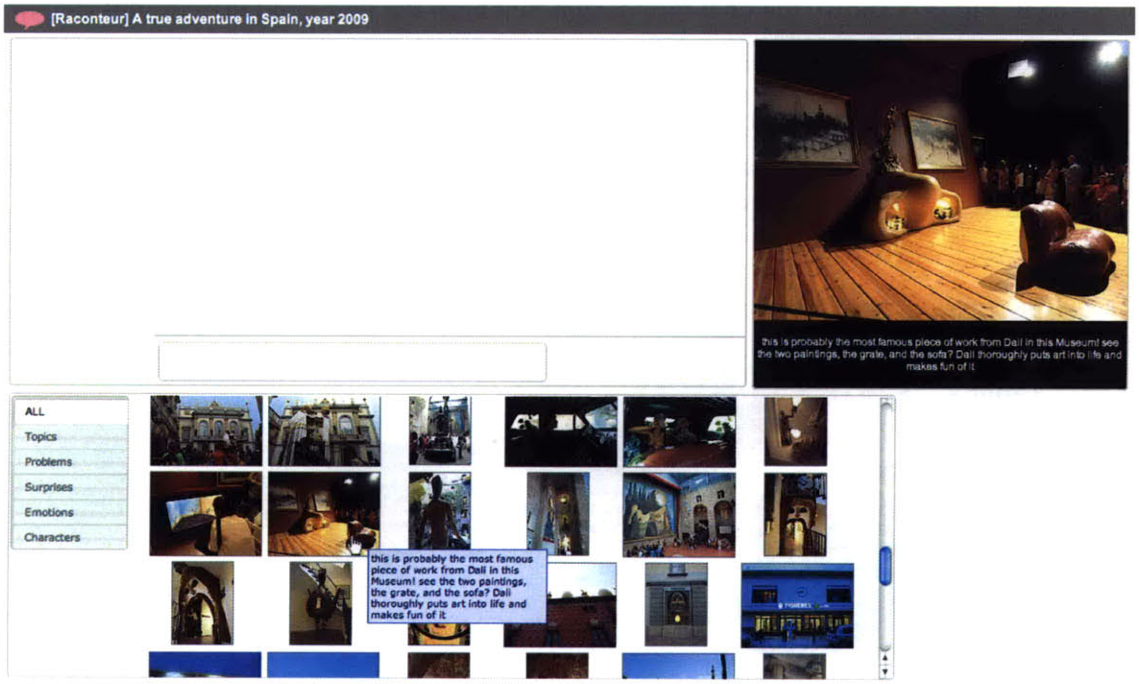

Fig. 5-3. The initial state of the Raconteur interface with a chat box, preview box, and the raw set of material fetched from the online album

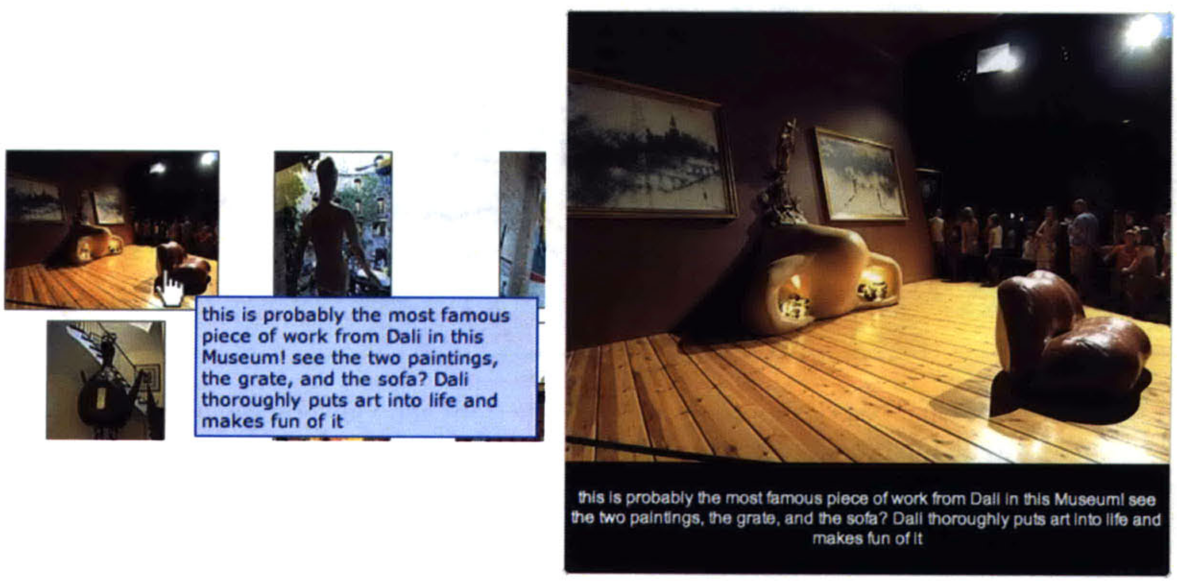

Fig. 5-4. User can preview the media elements by mouse hover.

5.2 Storytelling as Easy as Chat

When both users (a storyteller and a story viewer) log in to the system, users can start to chat. Fig. 5-5 shows the chat box design that both the teller and viewer will see. In this chat box, users can input narrations in plain text and send to the other user (Fig. 5-5a). They can see the messages from both users (Fig. 5-5b).

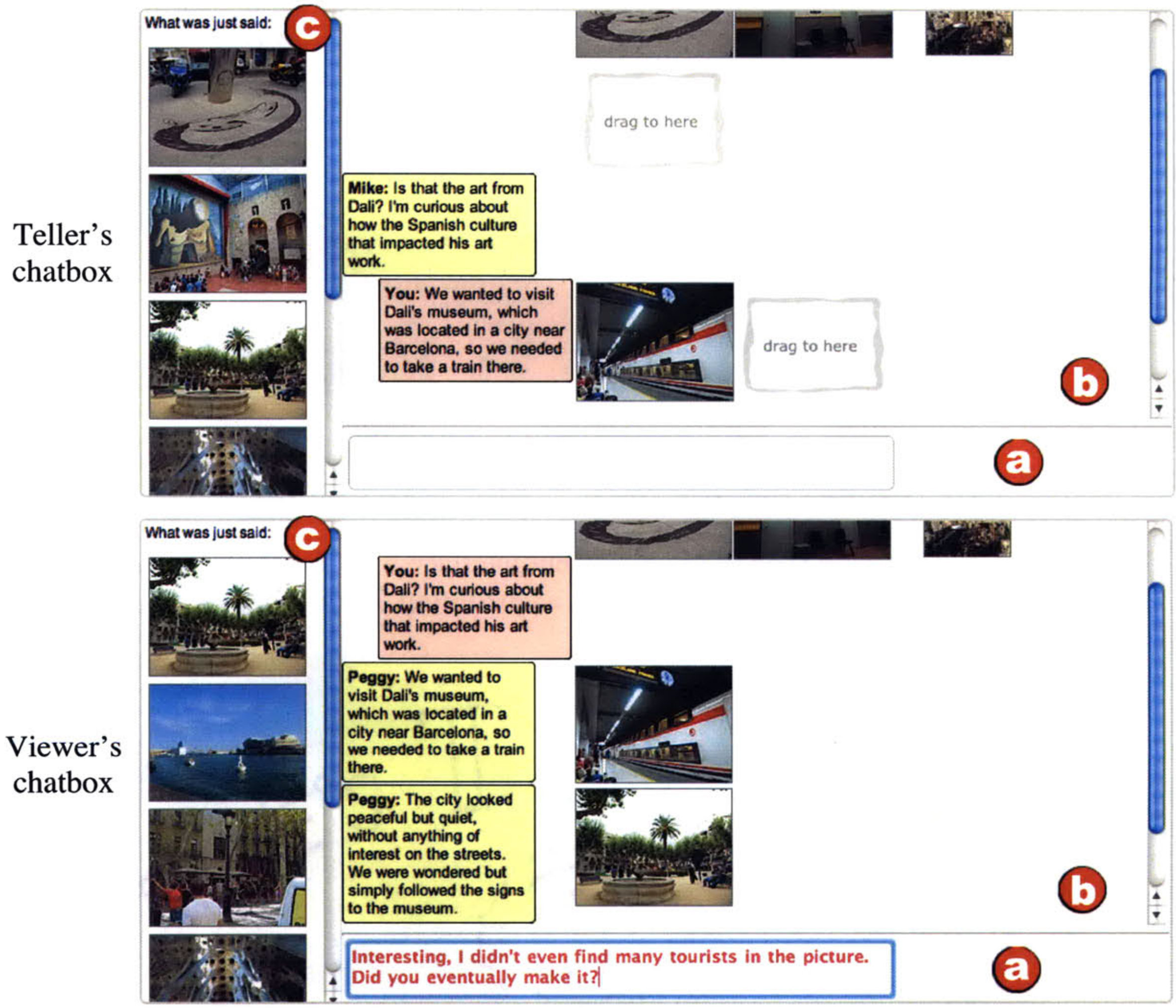

Fig. 5-5. Chat box design of the teller's side (top) and viewer's side (bottom), each including a) an input box, b) a list of chat messages and files, and edit areas for the storyteller, and c) Raconteur's found elements that match the latest narration.

In addition to using the chat box, users can also choose to chat about any media element in the system by mouse clicking an element and inputting a message into a pop-up speech bubble (shown in Fig. 5-6). In this way, users can directly illustrate the stories or comment on a specific photo or video.

Fig. 5-6. User can directly chat on any media element in the system. For example, the story viewer responds to the teller by posting a question specific to one of the edited photos.

5.3 Editing with Suggestions

When the storyteller sends out the chat message, Raconteur analyzes the narration and updates the interface to show the matched media elements (Fig. 5-5c). A storyteller can edit elements using the drag-and-drop interaction to attach files to the chat message in any desired order (Fig. 5-7). To help participants to reason about how the results were generated, we highlight the concept keywords that relate to system inferences. Moreover, when the user edits the elements, the system will interactively update the found patterns.

Fig. 5-7. Edit elements to enhance narrations by drag-and-drop.

5.4 Observing Story Patterns

Raconteur not only shows the matched elements to the narration, but also presents the relevant story patterns for the teller to consider how he or she wants to develop the story and continue the conversation. As shown in Fig. 5-8, the user can browse through different suggestions and story paths, such as several problems that happened during the trip and their story sequences. The teller can edit the files or chat on any element to go on the conversation with the viewer. Note that this panel that includes the whole media repository and the suggested patterns will not be shown to the viewer, to motivate him to follow the teller and explore the story as it unfolds.

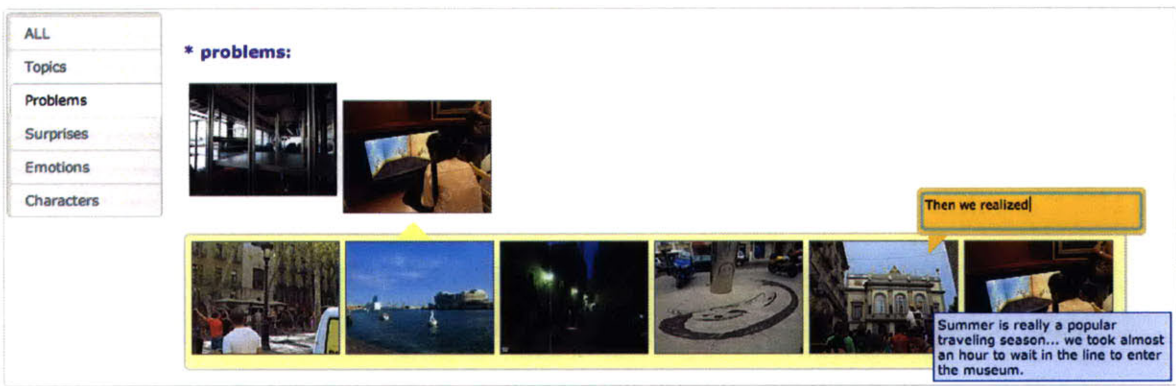

Fig. 5-8. Raconteur's suggestion panel of story patterns

Chapter 6

Evaluation and Discussion

This chapter presents the evaluation, results, and findings of inviting users to experience the Raconteur system.

6.1 User Study Goals

We conducted a usability study to answer the question: Does Raconteur succeed in providing assistance for conversational storytelling with personal digital media? The goal was to see:

- If users understood what Raconteur was for,
- Whether users chose to take advantage of Raconteur's assistance with storytelling, and when they did,
  - Whether they felt like Raconteur provided value in enhancing their storytelling or story listening experience, and
  - Whether Raconteur's interface was usable and enjoyable.

6.2 Participants and Material

We invited participants who were interested in sharing their personal stories with others. Each of the invited participants would take the role as a storyteller and invite another person he or she knew as a story viewer. This person in pair could be a friend, a family member, or in another relationship.

We conducted this study with 10 participants as 5 pairs (5 storytellers and 5 story viewers), of whom half were male and half were female, aged from 23-32 years old. These were different users than our formative user studies, so they were all new to the Raconteur system before the tests. They were all native or fluent speakers of English, without difficulties in reading or writing. All of them were frequent users of social network websites, with accounts in their own names. They updated their social network status once every four days on average, and updated personal albums with photos and/or videos once per week. Most enjoyed keeping their friends up-to-date about their activities, and in return, expected their friends to respond by adding comments, "thumbs-up" approval, forwarding, or reciprocal sharing.

We asked participants who served as storytellers to bring samples of their personal media files and orally tell the experiment facilitator their stories. The files could be from any media capture device, including a digital camera, camera phone, camcorder, or others. If the participant brought more than one set, the facilitator chose an appropriate set according to the complexity of the media collection, to avoid those that were too simple to provide any interesting feedback, or too complicated to fit within the allotted time. There was no constraint on the story topic.

Table 6-1 shows the 5 story topics chosen and the details of the collected material. Note that story sets #3 and #4 of similar topics were from distinct users with different main events and story characters taken at different times. The participants were asked to select the files from their own captured media sets and upload them to our Picasa test account. The story sets #2, #4, and #5 were originally also uploaded to Facebook for sharing and all had friends' comments. On average, the size of each uploaded repository was 60.8 media elements, containing 98.0% still photos and 2% short video clips (most within 30 seconds). 97.2% of the files were annotated; the average length of each caption was 10.0 English words. Each pair of the participants then would be asked to chat and edit the storyteller's material through the Raconteur system.

		media elements		annotation	
Story	# of main characters	# of photos	# of videos	# of files	average length
1) A 5-day sponsor visit to Italy	5	55 (96.5%)	2 (3.5%)	57 (100.0%)	11.1
2) A one-week trip to Spain for a conference demo and a proposal night	2	54 (98.2%)	1 (1.8%)	52 (94.6%)	12.2
3) A one-day beach party in summer 2009	8	51 (100.0%)	0 (0.0%)	51 (100.0%)	9.4
4) A one-day beach party in summer 2010	12	94 (98.9%)	1 (1.1%)	92 (96.8%)	9.0
5) A weekend at Pittsburgh for a friend gathering	6	90 (96.8%)	3 (3.2%)	89 (95.7%)	9.6
AVERAGE	6.2	68.8 (98.0%)	1.4 (2.0%)	68.2 (97.2%)	10.0

Table 6-1. Details of participants' uploaded media sets for the study

6.3 Procedure

The procedure of our evaluation was as follows:

1) We conducted a short pre-test interview to understand users' daily habits concerning media capture and editing, and to select a set of material to be used in the test.
2) For each set of material, we asked participants to annotate files with short captions in unrestricted English.
3) We introduced Raconteur and the interface with a 2-minute demonstration.
4) We conducted a storytelling session for each pair (a storyteller and a viewer) using Raconteur. In this session, a teller and a viewer were located in different rooms to avoid face-to-face communication. The users were allowed to chat and edit through the Raconteur interface until they decided to finish the conversation. We video recorded the storyteller's screen for later analysis.

5) We conducted a post-test interview for each pair together, to ask them to explain some of the decisions they had made, fill out a questionnaire, and provide comments, if any.

6.4 Measurement

To determine the effectiveness of the system, we quantitatively evaluate the following items:

- For the storytellers:
  - The total numbers of the edited files.
  - The source of the edited files from:
    - Raconteur's narration-matched list,
    - The suggestion panel of story patterns, and
    - The raw repository.

    This is to understand if the system's suggestions were effective and helpful to enhance the users' intentions or continue a topic rather than editing from a raw repository.
  - The percentage of the edited files by drag-and-drop and by click-and-chat.
  - How often they responded to the story viewer's questions or comments.
- For the story viewers:
  - How often they responded to the storyteller's points or edited files.
- For both the users:
  - The total numbers of the chatted messages,
  - The lengths (the word count) of each message, and
  - The results of the questionnaire using a Likert-5 scale.

Moreover, we qualitatively observed:

- The process of how story paths were developed,
- The degree of engagement of the chat process, and
- Any comments the participants wished to make during the chat.

6.5 Results and Discussion

We analyzed and summarized our collected results as below, including the quantitative analysis and qualitative findings.

6.5.1 Quantitative Results

The following tables show the quantitative results of this study, including the analysis of the five conducted chat sessions and the numbers of digital elements used in the chatted stories:

		# of chatted messages			*ave length of message*	
Story	total chat time (min)	total	from teller	from viewer	from teller	from viewer
1)	20	90	55 (61.1%)	35 (38.9%)	6.5	5.6
2)	25	132	59 (44.7%)	73 (55.3%)	7.0	5.5
3)	20	107	65 (60.7%)	42 (39.3%)	4.5	6.3
4)	25	134	63 (47.0%)	71 (53.0%)	7.4	4.7
5)	25	125	68 (54.4%)	57 (45.6%)	7.2	6.1
AVG	23	117.6	62 (52.7%)	55.6 (47.3%)	6.5	5.6

Table 6-2. The facts about the five chat sessions

			Interaction style		*Source of editing*		
Story	reposi-tory	edited files	by drag	by chat on file	narration match	pattern match	raw repository
1)	57	18 (31.6%)	12	6	14	4	0
2)	55	16 (29.1%)	9	7	11	3	1
3)	51	30 (58.8%)	21	9	24	6	0
4)	95	15 (15.8%)	14	0	11	4	1
5)	93	32 (34.4%)	22	10	24	8	0
AVG	70.2	22.2 (33.1%)	15.8 (71.2%)	6.4 (28.8%)	16.8 (75.7%)	5 (22.5%)	0.4 (1.8%)

Table 6-3. The analysis of the numbers of the media elements being edited by storytellers

The findings from the quantitative results were summarized as below:

#1: The length of chats

As shown in Table 6-2, the average time of a chat session was 23 minutes; the average chatted story contained 117.6 messages, 52.7% from storytellers and 47.3% from viewers. Note that one story point may be presented in several sequential messages, and one single event may also be divided into several messages, i.e. the numbers of messages do not indicate individual story events or topics. For example, a teller clicked on a photo and said, "*Check this out*." After sending it, he then continued explaining the sent photo, "*That shows how we "broke" the watermelon with a bat on the beach*."

Generally speaking, the conversations were balanced between the tellers and viewers, i.e. they chatted interactively instead of having one side dominating the conversation. Storytellers' chat messages were generally longer (6.5 words on average), while the viewer's messages were mostly short comments or questions (with an average of 5.6 words).

#2: The use of media elements

In the created stories, 33.1% of the media elements from tellers' repositories were used in a story (Table 6-3). There is no obvious relation between the size of repository and the number of used elements, i.e. a repository with a larger number of files does not imply a chat story with more edited elements.

As for the editing style, 71.2% of the edits were by dragging-and-dropping a Raconteur-suggested media element into the conversation. The users first narrated the stories with text messages, observed the matched elements, and then selected the files to enhance their narrated stories. 28.8% of the edited files were used via click-and-chat, i.e. users saw a media element and decided to talk about it by chatting on that element.

#3: The source of edited media elements

Table 6-3 listed the distribution of the source of the edited elements: 75.7% of the edited files were from direct narration match, 22.5% from Raconteur's suggestion panel with story patterns, and 1.8% from the raw repository. As for the categories of the used patterns, 80% were from the pattern of similar topics and 20% from the pattern of problems and resolutions, while categories of expectation violation, emotions, and characters were not used.

Question type	Details	Teller	Viewer
Understanding to Raconteur	Q1: "I understood what Raconteur was for and how to use Raconteur."	4.4	4.2
Easiness in general	Q2: "I found Raconteur easy to use."	4.4	4.8
Assistance in general	Q3: "Raconteur helped me tell my story." / "Raconteur helped me learn about my friend's story."	4.2	4.8
Assistance on continuing stories	Q4: "Raconteur supported me finding the elements I want." / "Raconteur enabled me to ask questions and explore the story as I want."	4.2	4.0
Assistance on story points	Q5: "Raconteur helped me make impressive "points" to my friend." / "Raconteur helped me remember the impressive "points" my friend made."	4.8	5.0
Information	Q6: "My created story was more informative than only reading my captions." / "The story was more informative than reading only captions."	4.2	4.8
Understanding to stories	Q7: "I believe my friend knows my story better through chatting with Raconteur." / "I know my friend's story better through chatting with Raconteur."	4.8	5.0
Easiness of editing & browsing	Q8: "It was easy to edit files while I was chatting." / It was easy to browse photos or videos while I was chatting.	3.6	4.2
Enjoyment	Q9: "I enjoyed using Raconteur."	4.6	4.8
Future opportunities	Q10: "I would use Raconteur again if I was telling a story." / "I would use Raconteur again if I was learning about a friend's story."	4.6	4.8

Table 6-4. Analysis of results from the questionnaire using Likert-5 scale, in which a 5 means "strongly agree" and 1 as "strongly disagree"

We discuss the potential reasons of the above analysis and the results from the questionnaire (Table 6-4) in the following session along with our observations.

6.5.2 Qualitative Results

Finding #1: Create stories as easily as in daily conversation.

All the participants "agreed" or "strongly agreed" Raconteur was easy to use (Q2 in Table 6-4). From the storytellers' point of view, the most intriguing aspect of the system was that they were able to transfer their comfort with the chatting process to a newfound comfort with the story composition process. One explained, "Talking to my friend and seeing Raconteur's suggestions helped me recall and brainstorm my stories. I was not thinking alone!" and another said, "In this process I was confident to talk about my stories, and I knew my friend was following so I could keep talking."

We also observed the chats were natural and similar to common dialogues in daily lives, where the storytellers first established the relations with the viewers to ask and talk about their recent status, and then started to share the experience with the captured material. Finally, they concluded by summarizing the trip or the activity, and said good-bye to each other to end the conversation. In the post-test interviews, they did not regard this process as video editing or composition, but as a chat that was enhanced by visual material. One storyteller explicitly explained the chat interaction was so natural that the system recommendation became "invisible" to him, i.e. he mainly focused on continuing the story with his friend instead of judging the system performance, as he often did when purposely searching images on the web.

Some users particularly liked the feature of commenting on a chosen file directly. When users clicked to chat about a media element (28.8% of the edits were from this interaction, see Table 6-3), storytellers often used conversational clauses such as "*this shows how* (…)", "*like this one*", "*check this out*", "*Did you see* (…)", "*have you seen this before?*", etc. to address the story viewers' attentions and make sure they were following. This observation is consistent in the essential elements of turn taking and recipient designing the stories of conversational storytelling as Polanyi [1989] explained (Chapter 2.2). One storyteller also explained that she recalled a topic they had chatted about before, face-to-face. During the chat, she became excited when she found the exact visual image to enhance their earlier conversation. It's also worth noting that this kind of situation is often marked by phrases such as "*by the way*", "*btw*", "*do you remember* (...)", etc.

All of the participants thought that chatting through Raconteur was "enjoyable" or "very enjoyable" (Q9 in Table 6-4 with average scores 4.6 and 4.8 from tellers and viewers respectively). During the chat sessions, we also observed the storytellers, the viewers, or both laughed or chuckled at many moments, such as making fun of themselves, explaining a file, or even seeing a surprising but matched search result. All of them were able to respond to most of their friends' questions or comments by explaining with more story details.

However, there were also issues participants noticed that we need to address: Chatting with a friend can sometimes be very intimate or go off topic because of the close relationships. The chat messages especially contain personal opinions, and conversational narrative is less structural for reading by outsiders. Moreover, some participants were not sure how well the system would work if they were confronted by an aggressive viewer who frequently interrupted. The nature of conversational storytelling makes this system mainly for personal, one-time enhanced chat. If the future system would be considered as a new video-editing interface, it might be helpful to track the relevance of the chat messages and to incorporate a phase of reviewing a created story before final publication.

Finding #2: Construct stories by connecting elements.

We were pleased to see that when editing elements, users followed Raconteur's suggestions about 98.2% of the time to construct stories and connect the events (from Table 6-3, 75.7% from the narration match and 22.5% from the suggested story patterns), instead of looking for files from the repository (1.8%). One participant said, "At first I thought it was more like real-time showing and commenting on my photos to my friend, but after seeing the suggested follow-up stories that illustrating my points, I soon realized I was connecting my experiences together." Another participant expressed, "Before the chat, I didn't have a clear structure in my mind how I should say something about my trip, but Raconteur's suggestions helped me put all these together and continue the topics. From my friend's response, I believe he understood my point and was engaged in my story."

On average the five participants agreed that Raconteur helped them tell the stories (Q3 in Table 6-4) and supported them finding the elements they wanted (Q4), both with an average score 4.2. Also, based on our observation of the created stories, storytellers were able to handle most of the conversation in a coherent fashion, avoiding abrupt, discontinuous jumps in topic.

Although the authors mostly edited from Raconteur's suggestions, there were also occasions that they accepted suggestions not because of the correctness, but because of the unexpectedness of the results. For example, in story #2 the teller said, "I remember seeing a giraffe figure that 'stood' on a porch waving happily," the system showed both the photo he was looking for and another one with a different subject, "This smiling wax figure of Einstein simply sat with all the staff at the front desk of the conference center..." (which includes the matched concepts: "figure" and "figure"; "stand" and "sit"; "wave", "happy" and "smile"). The teller laughed when he first saw it, and changed the topic to this after he edited the target file.

Most of the storytellers mentioned the update speed of system's suggestion was sometimes too fast. Instead of chatting on a file directly, at times they decided to input messages in the chat box and then drag the target file, but failed because the system had updated the list based on the typed narration. They suggested to us adding the previous/next-page function to browse the history of the matched element list. Moreover, for the 1.8% of the edited files that the storytellers directly used files from the unorganized repository instead of following the system suggestions, reasons include:

- The file they were looking for was not annotated,
- The viewer raised a relevant question as a branch of the original story path, or
- The teller decided to end the current topic and started another point.

We are planning to enable Raconteur to record these moments and learn from the user intention for further assistance. It was also worth noticing that only 22.5% of the edits was from the story pattern suggestions, mostly from the similar topics and some from problem and resolution. Storytellers explained that in order to keep the conversation, they did not want to spend time navigating and observing the results, but to quickly take a glance for a matched element or topic to go on.

Finding #3: Make impressive points during the chat.

From the questionnaire, the high scores to the two questions indicated Raconteur helped make impressive "points" (4.8 from tellers vs. 5.0 from viewers to Q5 in Table 6-4) and helped the viewers know the stories better through chatting (4.8 and 5.0 to Q7). In the post-test interviews when we asked the viewers to recall the chatted stories, they were all able to recount the exciting, impressive points that they had not expected, such as an interesting game, a special performance, something the friend had achieved, etc. Participants all agreed that the resulting stories were more informative than only reading the captions (4.2 and 4.8 to Q6).

In addition, the design also helped the storytellers to present their uniqueness. One user said, "I could reflect on my own opinions and thoughts much more than simply putting material together. In this system, I let my friend know more about what I have accomplished." Some selected examples from the tellers in the conversations include: "In the conference, my demo was a hot spot. I've even collected drawings from more than 80 participants. I was quite excited about this." and "It was really hard to resist the low temperature of the water, but that was not a problem to me as I often work out and swim." This aspect of the system is consonant with the view of life stories presented by Linde [1993].

Nevertheless, the turn-taking nature of a conversation also makes a created story less structural. Sometimes it was not so easy to see events in a clear chronological order, so in the post interview, some viewers explained they were not able to retell the friend's stories in a clear sequence when the storytellers brought up several topics in a short span.

Finding #4: High level of audience engagement in the stories.

All story viewers reported increased engagement in the story, particularly due to the reinforcement of the visual material and the real-time nature of the interaction. The post interviews showed the viewers could all remember and recall the story details. Participants said, "It was so impressive to see the pictures and understand the content when I was chatting."

and "I usually found myself getting lost after I watched a slideshow of an online album, but using Raconteur brought me into the scenes." Moreover, this interaction helped the audience achieve some degree of control of the story content: "I also could see how my friend chose the specific scenes based on my questions. I'm glad that my questions were heard and I could somehow control how the story could be developed." A few days after the test, one viewer even reported to us that he still talked about the story details with the teller in their face-to-face conversation when they were talking about another related topic. "I think this interaction has brought impact into my everyday life", he said.

Finding #5: Enrich media files through chatting.

We also observed that users often added new, or more complete, information to the media elements when they chatted about the story, instead of just repeating the annotations. The most obvious example was to explain the background of a character to complement the annotation (e.g. "The bass player tied a bell around his ankle so he was dancing all the time while performing." to the original caption "You will find street art performance everywhere, but this music band was especially incredible. We stopped here for many songs, and eventually bought several CDs of theirs." "The main organizer got into an accident before the opening but he still showed up." to "Milton came to say goodbye.").

Users also presented their goals and background information that might not easily be seen from captions (e.g. "The conference demo, which was the purpose of this trip, made me so nervous in the first few days. That was why I looked so worried, hardly with a smile, in those photos."). This showed that users were aware of the audience's story model, which is very important to compose a more accessible story. This also showed our approach would help the audience follow the content more smoothly compared to seeing a slideshow in chronological order.

Future Opportunities:

Considering a storyteller's conversation is adapted to different listeners with different tones and directions, we expect successive sessions with the same material will result in the system learning more information based on chat context and viewer preferences, which will enhance subsequent interactions. Several participants also expressed interest in a scenario where they could chat with other people with similar experiences, to compare with each other and collaborate creating a story together. Some participants discussed the use of the system in scenarios such as wedding parties and gathering amongst friends, in contrast to the travel scenarios tested. They characterized story topics in those scenarios as combining both the specifics of the particular events, and their past experiences.

Chapter 7

Related work

This section presents the related work to Raconteur in different research areas, including conversational interaction design with multimedia, media composition from natural language, and dialogue systems and models.

7.1 Chat, Collaboration, and Multimedia

Several research projects discuss the social media design and enrich the experience of collaboration or "chat" among several human users with multimedia data. Zync [Liu *et al.* 2007] is a plug-in video player to augment instant messaging software for social users to watch videos together and interact by chatting. Shamma *et al.* [2007] present an overview of different multimedia research approaches to utilize video content through studying online community activity such as collaborative viewing and chatting. Cesar *et al.* [2009] design a software architecture for media sharing across various users and devices with personalized content to enhance social interaction in a community. MapChat [Churchill *et al.* 2008] is a platform that enables users to chat on an interactive map and navigate the location-based information synchronously. Family Story Play [Raffle *et al.* 2010] is a device using video chat to support grandparents reading books together with young grandchildren.

The above projects focus on the chat interface to enable richer conversational experience with media instead of understanding the chat content between human users at a story level. Therefore, they differ from the goal of our research.

7.2 Media Composition using Natural Language

An emerging research area is to interact with digital media on the level of story composition in natural language. ARIA (Annotation and Retrieval Integration Agent) [Lieberman and Liu 2002] is a software agent that dynamically retrieves related photos based on the content of an email or web page. For example, when a user types his story of going to a friend's wedding, the system extracts the media annotation by considering roles such as who, what, where, and when, that address the story with similar context. Barry and Davenport [2003] presented a media capturing system that provides contextual information during the process of video capture in real-time to assist documentary. By reasoning the annotation of the current captured shot, it suggests the user (the videographer) making decisions such as what to record next, how to compose, and how to index the captured material at the phase of capturing raw material. ScriptSync is a feature for script-based video editing in the commercial software Media Composer [Avid Technology 2010]. Given a text-based script or transcript, it parses the content and phonetically associates the script with the source video clips that include spoken dialogues, i.e. it performs indexing by matching the sounds of human speech. This helps video editors compose from the story content in natrual languages, especially for the scenarios of interviews, documentaries, films, etc. Shen *et al.* [2009] designed a video editing system called Storied Navigation that enables authors to compose video clips, especially from a large documentary archive, by typing a story in unrestricted English sentences and retrieving relevant scenes. The system assists the authors focusing on the question "what's next?" of the current created story and deciding the continuing scenes.

These projects share the our goals of composing stories from annotated media clips and interacting with media elements considering story context; however, our work differs in several ways: 1) *The conversational interaction design for video-editing amateurs*: we focused on instantiating narrative goals directly through a chat scenario between two users, i.e. a storyteller and a viewer. Instead of putting a novice user into a system to capture or compose a set of material alone as the above systems, conversing with a friend helps the storyteller reason about those computational suggestions and motivates him to structure the story as daily conversation. 2) *The story editing design*: Raconteur's chat model also lets the storytellers "edit" media elements for showing friends instead of producing a final product. This makes the storytelling process more engaging in a higher level, in contrast to integrating individual files along a path. 3) *The use of the inference technique*: Raconteur uses the analogical inference tool AnalogySpace so that it not only determines concept similarity by one or a few relations, but from a more general aspect using vector computation with various features. This also helps the system identify larger story patterns among the elements, instead of one-time clip retrieval.

7.3 Dialogue Systems and Models

A dialogue system is a kind of computer system that interacts with a single user through conversations in various forms such as text, speech dialogues, and body gestures. It usually applies a dialogue model to define a coherent structure for the conversational interaction. For example, Stein *et al.* [1997] designed an intelligent multimedia retrieval system that helps user to clarify the information they want to access through a conversational process with a software agent. When a user makes a query "*Find 'Reichstag' after '1945'.*", the system reasons and responds with "*I can search for: 1. pictures; 2. biographies; 3. both.*" to interactively revise the search conditions and filter the results.

To converse with the user more naturally, some of the dialogue systems include virtual characters using a computer graphic or multi-model interface. Cassell [2001] presented research on the concept of an "embodied conversational agent" that represents an intelligent system as a virtual person to enable user experience similar to a face-to-face communication. AutoTutor [Graesser *et al.* 2001] is a tutoring system that helps students learn a subject through a conversation with an avatar with a talking head. Spierling and Iurgel [2003] designed a platform that helps artists to make a storytelling script for a human user to converse with virtual characters in an interactive play around the topic of art.

These systems showed how making a conversation helps a computer user navigate an interface better, but a predefined dialogue structure is different from our design of having two users talk and create stories without constraints.

Chapter 8
Conclusion and Future Work

We have presented Raconteur, a system for conversational storytelling that provides intelligent assistance in illustrating a story with photos and videos from an annotated media library. It performs natural language processing on a text chat between two or more participants, and recommends appropriate items from a personal media library to illustrate a story. We suggest that a Commonsense inference technique can identify larger scale story patterns and provide helpful assistance for users in real-time storytelling. Our user study shows that people find Raconteur's suggestions particularly helpful in continuing the story point and developing a coherent story path with the support of relevant media files.

Future work will focus on modeling the user storytelling dialogue, considering the system suggestions of the story patterns should be more tailored to the user intention and the purpose of the story. We also are redesigning the system to automatically learn from the created stories to support the storytellers' future chats with different viewers or a wider audience, and to enable collaborative storytelling to combine multiple multimedia libraries. We aim for providing a fun and productive environment for storytelling. Maybe it will help your friends become more interested in listening to your vacation stories, after all.

References

[Abbott 2002] ABBOTT, H. P. *The Cambridge Introduction to Narrative*. Cambridge University Press, 2002.

[Ames and Naaman 2007] AMES, M. AND NAAMAN, M. Why We Tag: Motivations for Annotation in Mobile and Online Media. In *Proceedings of CHI 2007: the 25th International Conference on Human Factors in Computing Systems*, ACM Press, San Jose, CA, USA, 2007.

[Avid Technology 2010] AVID TECHNOLOGY. *ScriptSync: Script-based Editing Option*. http://www.avid.com/scriptsync, accessed Aug. 15, 2010.

[Barry and Davenport 2003] BARRY, B. AND DAVENPORT, G. Documenting Life: Videography and Common Sense. In *Proceedings of ICME2003: the 2003 IEEE International Conference on Multimedia and Expo*, IEEE Press, Baltimore, MD, USA, 2003.

[Bird *et al.* 2008] BIRD, S. KLEIN, E, LOPER, E. AND BALDRIDGE, J. Multidisciplinary Instruction with the Natural Language Toolkit. In *Proceedings of TeachCL '08: the 3rd Workshop on Issues in Teaching Computational Linguistics*, Columbus, Ohio, USA, 2008.

[Black and Bower 1980] BLACK, J. B. AND BOWER G. H. Story Understanding as Problem-Solving. In *Poetics*, vol. 9, pp. 223-250, 1980.

[Black and Wilensky 1979] BLACK, J. B. AND WILENSKY, R. An Evaluation of Story Grammars. In *Cognitive Science*, vol. 3 (3), pp. 213-230, 1979.

[Cassell 2001] CASSELL, J. Embodied Conversational Agents: Representation and Intelligence in User Interfaces. In *AI Magazine*, vol. 22 (4), pp. 67-84, 2001.

[Cesar *et al.* 2009] CESAR, P., BULTERMAN, D., JANSEN, J., GEERTS, D., KNOCHE, H., AND SEAGER W. Fragment, Tag, Enrich, and Send: Enhancing Social Sharing of Video. In *TOMCCAP: Transactions on Multimedia Computing, Communications, and Applications*, vol. 5 (3), 2009.

[Cheong *et al.* 2008] CHEONG, Y., KIM, Y. MIN, W. AND SHIM, E. PRISM: A Framework for Authoring Interactive Narratives. In *Proceedings of ICIDS 2008: the 1st Joint International Conference on Interactive Digital Storytelling: Interactive Storytelling*, ACM Press, Erfurt, German, 2008.

[Chi and Lieberman 2010] CHI, P.-Y. AND LIEBERMAN, H. Raconteur: From Intent to Stories. In *Proceedings of IUI 2010: the 14th International Conference on Intelligent User Interfaces*, ACM Press, Hong Kong, China, 2010.

[Churchill *et al.* 2008] CHURCHILL, E., GOODMAN, E., AND O'SULLIVAN, J. Mapchat: Conversing in Place. In *Extended Abstract of CHI 2008: the 26th International Conference on Human Factors in Computing Systems*, ACM Press, Florence, Italy, 2008.

[Cooper *et al.* 2003] COOPER, M., FOOTE, J., GIRGENSOHN, A., AND WILCOX, L. Temporal Event Clustering for Digital Photo Collections. In *Proceedings of MM'03: the 11th ACM International Conference on Multimedia*, ACM Press, Berkeley, CA, USA, 2003.

[Engström *et al.* 2010] ENGSTRÖM, A., JUHLIN, O., PERRY, M., AND BROTH, M. Temporal Hybridity: Footage with Instant Replay in Real Time. In *Proceedings of CHI 2010: the 28th International Conference on Human Factors in Computing Systems*, ACM Press, Atlanta, GA, USA, 2010.

[Fellbaum 1998] FELLBAUM, C. *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA, 1998.

[Garfinkel 1967] GARFINKEL, H. Common Sense Knowledge of Social Structures: The Documentary Method of Interpretation in Lay and Professional Fact Finding. In *Studies in Ethnomethodology*, Polity Press. pp. 76-103, 1967.

[Gentner 1998] GENTNER, D. Analogy. In *A Companion to Cognitive Science*, Oxford University Press, pp. 107-113, 1998.

[Gervas 2009] GERVAS, P. Computational Approaches to Storytelling and Creativity. In *AI Magazine*, vol. 30 (3), pp. 49-62, 2009.

[Graesser *et al.* 2001] GRAESSER, A.C., VANLEHN, K., ROSÉ, C.P., AND JORDAN, P.W. Intelligent Tutoring Systems with Conversational Dialogue. In *AI Magazine*, vol. 22 (4), pp. 39-51, 2001.

[Harrell 2006] HARRELL, D. F. Walking Blues Changes Undersea: with the GRIOT System Imaginative Narrative in Interactive Poetry Generation. In *Proceeding of the AAAI 2006 Workshop in Computational Aesthetics*, AAAI Press, Menlo Park, CA, USA, 2006.

[Harrell and Chow 2009] HARRELL, D. F. AND KENNY K. N. CHOW. Generative Visual Renku: Linked Poetry Generation with the GRIOT System. In *Hyperrhiz: New Media Cultures*, vol. 6, 2009. Available from: http://www.hyperrhiz.net/hyperrhiz06/19-essays/79-generative-visual-renku.

[Havasi *et al.* 2007] HAVASI, C., SPEER, R., AND ALONSO, J. ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge. In *Proceedings of RANLP 2007: Recent Advances in Natural Languge Processing 2007*, Borovets, Bulgaria, 2007.

[Havasi 2009] HAVASI, C. *Discovering Semantic Relations Using Singular Value Decomposition Based Techniques*. PhD thesis. Brandeis University, p. 43, 2009.

[Joshi and Luo 2008] JOSHI, D. AND LUO, J. Inferring Generic Activities and Events from Image Content and Bags of Geo-tags. In *Proceedings of CIVR'08: the 2008 International Conference on Content-based Image and Video Retrieval*, ACM Press, Niagara Falls, Canada, 2008.

[Kirk *et al.* 2006] KIRK, D., SELLEN, A., ROTHER, C., AND WOOD, K. Understanding Photowork. In *Proceedings of CHI 2006: the 24th International Conference on Human Factors in Computing Systems*, ACM Press, Montréal, Québec, Canada, 2006.

[Kuipers 1984] KUIPERS, B. Commonsense Reasoning About Causality: Deriving Behavior from Structure. In *Artificial Intelligence*, vol. 24, pp. 169-203, 1984.

[Labov and Waletzky 1967] LABOV, W. AND WALETZKY, J. Narrative Analysis: Oral Versions of Personal Experience. In *Essays on the Verbal and Visual Arts*, University of Washington Press, pp. 12-44, 1967.

[Labov 1997] LABOV, W. Some Further Steps in Narrative Analysis. In *Journal of Narrative and Life History*, vol. 7 (1-4), pp. 395-415, 1997. Available from: http://www.ling.upenn.edu/~wlabov/sfs.html

[Lebowitz 1985] LEBOWITZ, M. Story Telling as Planning and Learning. In *Poetics*, vol. 14, pp. 483-502, 1985.

[Lieberman and Liu 2002] LIEBERMAN, H. AND LIU, H. Adaptive Linking Between Text and Photos using Common Sense Reasoning. In *Proceedings of AH2002: the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-based Systems*, ACM Press, London, UK, 2002.

[Linde 1993] LINDE, C. *Life Stories: The Creation of Coherence*, Oxford University Press, 1993.

[Liu and Singh 2004] LIU, H. AND SINGH, P. ConceptNet: a Practical Commonsense Reasoning Toolkit. In *BT Technology Journal*, vol 22 (4), pp. 211-226, 2004.

[Liu *et al.* 2007] LIU Y., SHAFTON, P. SHAMMA, D., AND YANG, J. Zync: the Design of Synchronized Video Sharing. In *Proceedings of DUX '07: the 2007 Conference on Designing for User Experiences*, ACM Press, Chicago, IL, 2007.

[Mateas and Stern 2003] MATEAS, M. AND STERN, A. Façade: an Experiment in Building a Fully-realized Interactive Drama. In *Game Developer's Conference: Game Design Track*, 2003.

[Montfort 2009] MONTFORT, N. Curveship: an Interactive Fiction System for Interactive Narrating. In *Proceedings of CALC '09: the Workshop on Computational Approaches to Linguistic Creativity*, pp. 55-62, 2009.

[Mueller 2006] MUELLER, E. T. *Commonsense Reasoning*, Morgan Kaufman, 2006.

[Polanyi 1989] POLANYI, L. *Telling the American Story: A Structural and Cultural Analysis of Conversational Storytelling*, MIT Press, 1989.

[Prince 2003] PRINCE, G. *A Dictionary of Narratology,* University of Nebraska Press, 2003.

[Raffle *et al.* 2010] RAFFLE, H., BALLAGAS, R., REVELLE, G., HORII, H., FOLLMER, S., GO, J., REARDON, E., MORI, K., KAYE, J. AND SPASOJEVIC M. Family Story Play: Reading with Young Children (and Elmo) Over a Distance. In *Proceedings of CHI 2010: the 28th International Conference on Human Factors in Computing Systems*, ACM Press, Atlanta, GA, USA, 2010.

[Riedl and Leon 2009] RIEDL, M. AND LEÓN, C. Generating Story Analogues. In *Proceedings of AIIDE '09: the 5th International Conference on Artificial Intelligence for Interactive Digital Entertainment*, AAAI Press, Palo Alto, CA, USA, 2009.

[Schank *et al.* 1982] SCHANK, R. C. AND G. COLLINS, E. DAVIS, P. JOHNSON, S. LYTINEN, B. REISER. What's the Point? In *Cognitive Science,* vol. 6, pp. 255-275, 1982.

[Schank 1986] SCHANK, R. C. *Explanation Patterns: Understanding Mechanically and Creatively*, Psychology Press, 1986.

[Schank 1991] SCHANK, R. C. *Tell Me a Story: A New Look at Real and Artificial Intelligence*, Northwestern University Press, 1991.

[Shamma *et al.* 2007] SHAMMA, D., SHAW, R., SHAFTON, P, AND LIU, Y. Watch What I Watch: Using Community Activity to Understand Content. In *Proceedings of MIR '07:*

the International Workshop on Multimedia Information Retrieval, Ausburg, Bavaria, German, 2007.

[Shen *et al.* 2009] SHEN, Y.-T., LIEBERMAN, H., AND DAVENPORT G. What's Next? Emergent Storytelling from Video Collections. In *Proceedings of CHI2009: the 27th International Conference on Human Factors in Computing Systems*, ACM Press, Boston, MA, USA, 2009.

[Speer *et al.* 2008] SPEER, R., HAVASI, C., AND LIEBERMAN, H. AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge. In *Proceedings of AAAI2008: the 23rd AAAI Conference on Artificial intelligence*, AAAI Press, Chicago, IL, USA, 2008.

[Spierling and Iurgel 2003] SPIERLING, U. AND IURGEL, I."Just Talking about Art"-Creating Virtual Storytelling Experiences in Mixed Reality. In *Proceedings of ICVS '03: the 2nd International Conference on Virtual Storytelling*, Toulouse, France, 2003.

[Stein *et al.* 1997] STEIN, A., GULLA, J.A., MÜLLER, A., AND THIEL, U. Conversational Interaction for Semantic Access to Multimedia Information. In *Intelligent Multimedia Information Retrieval*, MIT Press, pp. 399-421, 1997.