Intelligent Assistance for Conversational Storytelling Using Story Patterns

Pei-Yu (Peggy) Chi
MIT Media Lab
20 Ames St.
Cambridge, MA, USA
peggychi@media.mit.edu

Henry Lieberman
MIT Media Lab & MIT Mind-Machine Project
20 Ames St.
Cambridge, MA, USA
lieber@media.mit.edu

ABSTRACT
People who are not professional storytellers usually have difficulty composing travel photos and videos from a mundane slideshow into a coherent and engaging story, even when it is about their own experiences. However, consider putting the same person in a conversation with a friend – suddenly the story comes alive.

We present *Raconteur 2*, a system for conversational storytelling that encourages people to make coherent points, by instantiating large-scale story patterns and suggesting illustrative media. It performs natural language processing in real-time on a text chat between a storyteller and a viewer and recommends appropriate media items from a library. Each item is annotated with one or a few sentences in unrestricted English. A large commonsense knowledge base and a novel commonsense inference technique are used to identify story patterns such as *problem and resolution* or *expectation violation.* It uses a concept vector representation that goes beyond keyword matching or word co-occurrence based techniques. A small experiment shows that people find Raconteur's interaction design engaging, and suggestions helpful for real-time storytelling.

Author Keywords
Storytelling, conversation, chat, life stories, story pattern, commonsense computing, digital media, video, photograph.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors, Languages, Design

INTRODUCTION
Sharing life stories is a foundation of interpersonal communication. In this digital age, it is common to see a friend posting a series of photos and video clips through the online social platforms such as Facebook, Flickr, and YouTube. Usually, such a personal multimedia repository is full of individual media elements that include various story events [1]. However, for something like a two-week vacation, the sheer number of media elements may grow large. The experience may have many dimensions, and there are many possibilities for how to tell the story. Professional storytellers are good at connecting events and making interesting points, but amateur storytellers may not have the skill to compose an illustrated story that is engaging for the audience. As a result, most personal media collections are presented in chronological order [12], giving the stories a rather "flat" feel.

However, if we put the same storyteller into a face-to-face conversation with a sympathetic, interested and questioning listener, suddenly *the story comes alive*. Conversational storytelling is one of the most basic, familiar forms of communication in everyday life [21]. It involves at least one speaker and one listener that create stories together. For the speaker, it is as easy as making casual conversation, but it has a greater purpose – to share life stories, composed of important narrative elements such as characters, events, and causal connections. For the listener, the aim is to respond to, and to acknowledge what has been said, as well as give feedback and perhaps reciprocate by sharing his or her own stories. Story structure emerges through such interaction, and the participants come to understand the specific context, and communicate more effectively.

Studies have shown there are thriving conversations between online users over this user-generated content through weblogs and social networking websites [4,14]. Users not only share personal multimedia, but also associate contextual information such as adding captions, changing titles, making comments, etc. At the same time, the audience responds with their own comments or similar personal stories, which motivate the authors to tell more about the experience. In other words, people *chat about life stories through digital media* to enhance their personal relationships. This user interaction provides the opportunity for intelligent systems to understand the narrative intent behind the conversation, and make suggestions for relevant media elements.

IUI 2011, February 13–16, 2011, Palo Alto, California, USA.

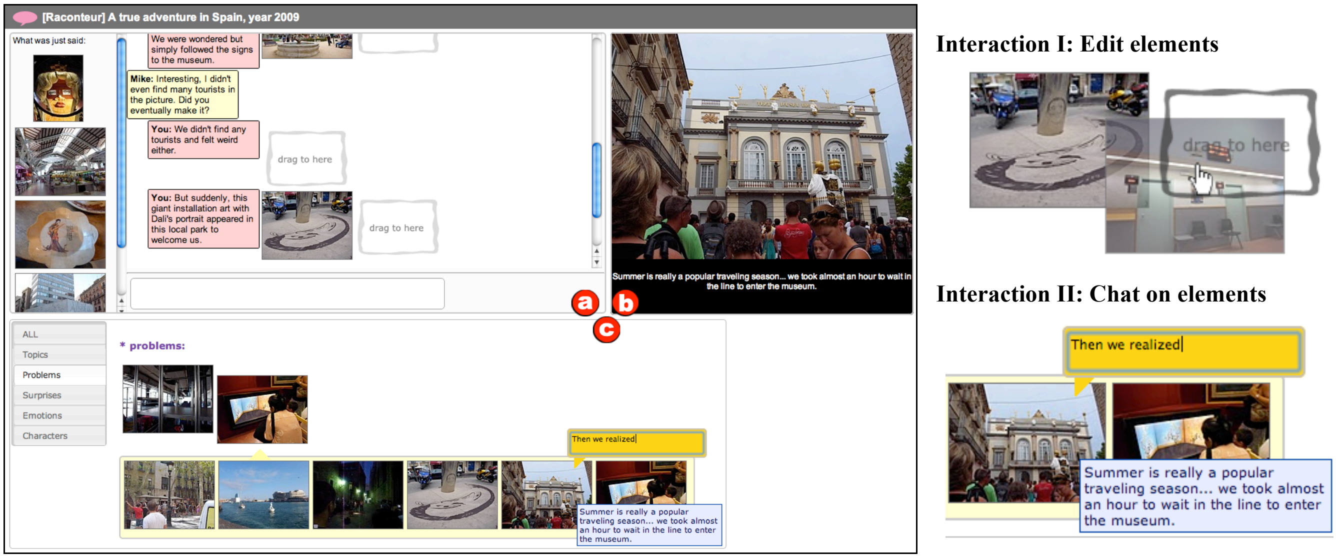

Fig. 1. Raconteur user interface, including: a) a chat box where storytellers can chat with a friend in text, see the matched media elements, and edit to enhance his story, b) a preview window to see the photos and videos with captions, and c) Raconteur's suggestion panel for observing the story patterns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

We present *Raconteur 2* (henceforth, simply Raconteur), a personal story editing system that helps users think about story development in multimedia material by enabling conversations with friends – 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 patterns and paths for the storyteller to continue. Using natural language processing and Commonsense reasoning, Raconteur analyzes the multimedia items in a repository, each optionally annotated with textual information.

This paper makes the following contributions: Our focus on story pattern analysis shows how an intelligent system can make recommendations that go beyond simple keyword search or topic classification. Tracking the conversation over time can aid story development beyond a single search-and-retrieval interaction. Raconteur's role is not to generate stories by itself, nor to make communication decisions for the human user. Our interface design focuses on enabling human-human interaction through natural dialogue, with our system playing an assistive role.

CHATTING THROUGH RACONTEUR

Fig. 1 shows Raconteur's web-based user interface, where the *storyteller* is able to:

- Chat in plain text with a *story viewer*, a friend with whom he would like to share the experience (Fig. 1a bottom), see the matched media elements (Fig. 1a left), and preview the photos and videos with captions (Fig. 1b).
- Enhance his chatted story (Fig. 1a right) by drag-and-drop of media elements or directly chatting on any element directly.
- See Raconteur's suggestion panel (Fig. 1c), including the story patterns and the raw material of the repository.

On the other side, the story viewer will see the same interface (Fig. 1a and Fig. 1b) without the whole media repository or pattern suggestions. This will motivate the viewer to follow the teller and remain engaged, without the temptation to independently browse the media. 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 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 of a storyteller sharing a travel story with a friend as a story viewer (where T stands for the *Teller*, R as *Raconteur*, and V as the *Viewer*):

T: [Input in the chat box] "*My trip to Spain was full of surprising stories.*" (TellerMsg#1)

R: [Suggests these story points: 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]

T: [Select three suggested topics and drag the photos to TellerMsg#1]

V: [Click to chat about one of the photos] "*Is that the art by Dali? Tell me more about the visit! I'm curious about how the Spanish culture impacted his work.*" (ViewerMsg#1)

R: [Updates suggestions of the precedent and following elements of this story point]

In our storytelling model, Raconteur enables the storyteller to start to talk about his stories without constraints. The system understands the concepts of “trip”, “surprising”, and “stories” in the user narration. It reasons about the correspondence between the narrative goals and the concrete annotation. The storyteller selects from the system suggestions he would like to share by attaching photos or videos to his chat message. The viewer then chooses one topic and responds by raising a question. Raconteur matches his message to the story topic, and suggests the other media elements about this theme, to assist the teller in developing a particular story point.

T: [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.*” (TellerMsg#2)

R: [Suggests elements by the pattern: a photo taken outside of the station, and a photo of the installation art]

T: [Click to chat about the city view] “*The city looked peaceful but quiet, without anything of interest on the streets. We were wondered* (sic) *but simply followed the signs to the museum.*” (TellerMsg#3)
[Drag one more photo of the city view to TellerMsg#3]

R: [Suggests elements by the pattern: 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]

V: [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?*” (ViewerMsg#2)

T: [Click to chat about 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.*” (TellerMsg#4)

V: [Input in the chat box] “*Wow! Now you can be sure you have come to the right place to see Dali’s masterpiece!*” (ViewerMsg#3)

Based on the selected elements, Raconteur suggests the possible pattern of *expectation violation*. The teller chooses to continue describing his visit to the museum, explaining his first impression of the city. The viewer follows what the teller shares and finds the experience does not meet his expectations. Finally, Raconteur helps the teller, step by step, to make the point of “surprising stories” and create a story path that reflects both users’ interests.

Constructing story patterns is important for story understanding and sharing. Polanyi defined stories as “*specific past time narratives with a point*” [21]. 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 anything particularly remarkable, lest it become difficult to be remembered, retold, and therefore “dreary” [17]. Polanyi also explained “turn-taking” between participants, unlike speeches or interviews, and indicated that conversational storytellers “*are under a very strong constraint to make their utterances somehow coherent with what has been going on immediately preceding their talking*.”

Furthermore, to continue the stories in a conversation, it is important to structure personal stories so that a listener can reason about them. Schank proposed the idea of “story skeleton” to explain how we construct and comprehend a story [24]. He also suggested how the underlying story structure might alter the listening experience: “*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.*” Similarly, some researchers have addressed the concept of “story grammar” to support story composition by a set of rules [2]. Labov and Waletzky analyzed the structure of oral narrative of personal experience [16]. Their overall structure includes: orientation, complication, evaluation, resolution, and coda. The challenge to design an intelligent system lies on finding possible patterns to illustrate a story point.

RACONTEUR DESIGN

We designed the system to reason about stories from a personal multimedia repository for users to interactively chat and edit. Raconteur’s system structure is composed of several major components as follows:

- A *multimedia database* of media elements, annotated with textual information,
- A *narration processor* that parses the user’s narrations and captions,
- An *story reasoning model* that reasons about user narration and finds patterns using a Commonsense knowledge base, and
- A *user interface* that allows a pair of users (a *storyteller*, who owns the multimedia data, and his/her friend, the *story viewer*) to chat about the story, observe the system’s suggestions, and edit in real-time.

Multimedia Resources

For a given multimedia repository, we see each photo, video clip, audio file, or other media, all as individual “media elements”, i.e. story units in the system. Each of these elements can be annotated with a sentence or two in unrestricted natural language, as online users already often do. The annotation may describe characters, events that happened, and intent or opinion of the captured scene. For example, “*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 such descriptive information rather than simply subjects, objects, actions, etc., in order to acquire contextual relationships. There is also the possibility that annotations may be generated by metadata, transcription of audio, image recognition or other means, but for this prototype we only use hand-generated annotations. Any unannotated files will be kept in the repository, but not considered by the analysis. However, they can be referred to and attached if users explicitly specify. The narration that occurs during a chat may be considered as additional information for future reference, thereby enriching the media library for future chat episodes.

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 an Application Programming Interfaces (API) such as Picasa[1]. The Raconteur system needs to access users' album lists, titles, dates, descriptions, and the lists of files (photos and/or videos), each with file system links or hyperlinks of thumbnails and content, and their captions, file types, dates, etc.

Narration Processing

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

We applied the Natural Language Toolkit (NLTK) [3], a suite of programming libraries for symbolic and statistical NLP. We particularly use several features:

- Part of speech (POS) tagging to identify words including verbs, nouns, and adjectives/adverbs, 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.
- 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.
- Stemming and lemmatization to normalize words into the basic forms (e.g. "went" into "go", "the cars" into "car"), for later concept processing and comparison.

In addition to identifying interjections or reinitiation markers by NLP, we remove those non-story-world clauses that contain verbs, but do not provide story-related information, such as "think", "mean", "know", "guess", etc. based on literature study results [21].

Reasoning About Life Stories Using Commonsense

To reason about the events in the repository, we apply common sense knowledge, which 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, it has been long studied by the social sciences. The sociologist Garfinkel explains how common sense helps people interpret each other [10]: "… *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.*"

Background of Commonsense Computing

To enable computers to understand our stories and "think" more like humans, we need to help computers acquire this type of knowledge. Since 1999, researchers have been collecting common sense knowledge from volunteers on the Internet to build a knowledge base called Open Mind Common Sense (OMCS) in the form of 20 or so kinds of two-place relations [26]. For example, "*AtLocation*(art, museum)", means "Something you find at a museum is art.". "*PartOf*(sculpture, art)" can express "Sculpture is a kind of art.". Catchalls like "*HasProperty*(art, inspiring)", can express "Art is inspiring", even when we don't have *Inspiring* explicitly as a relation. Currently, the knowledge base in English has over a million assertions from over 15,000 contributors. This collected data is then represented by ConceptNet as a semantic network [19].

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. AnalogySpace is a powerful tool for analogical reasoning [27]. It 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. Inference is performed by Principal Component Analysis on this matrix, using the linear algebra factorization method called "Singular Value Decomposition" (SVD) to find axes which best characterize the data. These axes are often semantically meaningful, and enable us to measure abstract concepts quantitatively by vector calculation, i.e. making the abstract concepts computable. 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:

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

- Getting an ad-hoc category of a concept (e.g. “art”, “museum”, “sculpture” may fall into one category along with “painting” and “artist”),
- 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?”).

In this way, we can provide users the freedom of describing their stories without constraining their expression. In addition, we can also reason about the narration and understand the inferred intentions. This moves the system from word matching to story understanding, and most important of all, assisting storytelling.

Concept Vector Calculation

Based on the result of NLP, we traverse each verb, noun, adjective, and adverb 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 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 the previous example, the narration 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}).

Media Elements Association

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 similarity by a concept vector calculation containing the story elements.

The simplest measurement is to compare all the concepts of the annotations placed on two elements. We compare two annotations using their “narration vectors”: 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.

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 [9] 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 [13] and other connections, such as “buy” and “wallet”, or “beautiful” and “painting.”

Story Pattern Finding

To identify larger story patterns, we developed an inference technique that considers several patterns, which are structures that make similar points. Telling stories by making such enhanced points usually helps story listeners to understand and follow the story better. Each story path may provide a different story experience to the audience. Therefore, our goal is to find the elements with *connected events and similar intentions*.

Problem and Resolution

Based on our formative study [7], we found the most common pattern of travel stories is encountering unexpected problems. This often makes a personal story “special” and impressive to the audience because it arouses the listeners’ curiosity or reminds them of similar life experiences. We analyzed each annotation according to our concepts of *intention, problem, resolution, and consequence*. Examples include: the story “one-week trip to Spain” contains “buy fresh goods in a local market” (*intention*), “my 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 assembling the tent poles” (*problem*), “reading instructions” (*resolution*), and “successfully settling down together” (*consequence*).

We found that a good way to recognize mention of “problems” is to look for concepts that ConceptNet knows people “don’t like”. To detect this kind of concept, we reason using AnalogySpace: from the conjugate matrix 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, such as “delay” (-0.99), “traffic jam” (-0.99), “wait” (-0.24), “steal” (-0.03), “lose” (-0.11), etc., compared to other positive concepts that people like, such as "travel" (0.02), "famous" (0.69), "sunshine" (0.70), etc. This inference enables us to identify those possible problems in a repository. After identifying the potential problems happening in the stories, we then reason about the connected events related to each problem. These events can include causality relations, or simply around the same topics or with the same subjects (Fig. 2).

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. 2. Finding a collection of problems and resolutions

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". For 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 [23]. 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.*"

Similar Topics

Continuing a story with connected topics helps an audience formulate a coherent perspective on a story. For example, when talking about a conference or meeting, similar ideas such as organizers, presentations, posters, audience, etc. are often addressed. 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 the repository by associating the elements with each other.

Emotions and Characters

In addition, we identify several common types of emotion, using AnalogySpace vectors (v_{happy}, v_{relaxed}, v_{excited}, v_{worried}). Last but not least, Raconteur is also able to identify characters and locations by named entity recognition. Examples of named entities include human names and geographic names. However, we track not only known names, but also 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 directions*" and "*Our group photo with the famous landmark*".

Suggestion Updating in Real-time

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 to the interface. It detects the user edits that match to the paths, and avoids frequent suggestion of the same elements that have been edited and shown.

EVALUATION AND DISCUSSION

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.

Participants and Collected 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 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. 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" ("likes") approval, forwarding, or reciprocal sharing.

We asked participants who served as storytellers to bring samples of their personal media files and verbally 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.

Procedure and Measurement

The procedure of our evaluation was as follows: 1) We conducted a short pre-test interview to understand storytellers' 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 select the files from their own captured media sets, upload them to our Picasa test account, and 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.

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

- The numbers of the chatted messages and edited files,
- The source of the edited files from: Raconteur's narration-matched list, the suggestion panel of story patterns, and the raw repository,
- The numbers of the edited files by different interaction styles including drag-and-drop and click-and-chat, and
- The results of the questionnaire using a Likert-5 scale.

Results and Discussion

Facts About the Chats

Table 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 tellers with different main events and story characters taken at different times. 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 70.2 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.

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

	Media files		*Chatted messages*		*Interaction style*		*Source of editing*		
Story	reposi-tory	edited files	from teller	from viewer	by drag	by chat on file	narration match	pattern match	raw set
1) A 5-day sponsor visit to Italy	57	18 (31.6%)	55 (61.1%)	35 (38.9%)	12	6	14	4	0
2) A one-week trip to Spain for a conference	55	16 (29.1%)	59 (44.7%)	73 (55.3%)	9	7	11	3	1
3) A one-day beach party, 2009	51	30 (58.8%)	65 (60.7%)	42 (39.3%)	21	9	24	6	0
4) A one-day beach party, 2010	95	15 (15.8%)	63 (47.0%)	71 (53.0%)	14	0	11	4	1
5) A weekend at Pittsburgh for a friend gathering	93	32 (34.4%)	68 (54.4%)	57 (45.6%)	22	10	24	8	0
AVERAGE	70.2	22.2 (33.1%)	62 (52.7%)	55.6 (47.3%)	15.8 (71.2%)	6.4 (28.8%)	16.8 (75.7%)	5 (22.5%)	0.4 (1.8%)

Table 1. The analysis of participants' chats and edits with uploaded media sets for the study

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).

In the created stories, 33.1% of the media elements from tellers' repositories were used in a story. 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 storytellers 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. tellers saw a media element and decided to talk about it by chatting on that element.

75.7% of the edited files were from 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 explicitly used but were considered from the narration match list.

Create Stories as Easily as in Daily Conversation.

All the participants "agreed" or "strongly agreed" that Raconteur was easy to use. 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."

Construct Stories by Connecting Elements.

We were pleased to see that when editing elements, storytellers followed Raconteur's suggestions about 98.2% of the time to construct stories and connect the events (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."

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.

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) and helped the viewers understand the stories better through chatting (4.8 and 5.0 respectively). 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 respectively).

In addition, the design also helped the storytellers to present their uniqueness. One teller 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 from literature criticism [18].

Nevertheless, the turn-taking nature of a conversation also makes a created story less structural for reviewing afterwards. 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.

High Level of Audience Engagement in the Stories.

All story viewers reported high 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.

RELATED WORK

Our previous paper [7] reported a formative user study, and briefly described an earlier system, *Raconteur 1*, which concentrated on finding analogies between story elements. That system was designed for a single-user scenario, authoring media to be presented later. The present paper is aimed at a very different scenario, conversational storytelling in a real-time context. It concentrates on instantiating larger scale story patterns, and features a completely redesigned interface, emphasizing the conversational aspects of the chat interaction. The recommendation algorithm has been redesigned, and is explained here in full detail. The user study is also entirely new.

Several research projects discuss the social media design and enrich the experience of collaboration or "chat" among several human users with multimedia data. Zync is a plug-in video player to augment instant messaging software for social users to watch videos together and interact by chatting [20]. Shamma *et al.* present an overview of different multimedia research approaches to utilize video content through studying online community activity such as collaborative viewing and chatting [25]. Comic Chat is a system that enhanced online communication in the form of a comic strip with graphical representation in real-time [15]. Cesar *et al.* design a software architecture for media sharing across various users and devices with personalized content to enhance social interaction in a community [6]. MapChat [8] is a platform that enables users to chat on an interactive map and navigate the location-based information synchronously. Family Story Play [22] is a device using video chat to support grandparents reading books together with young grandchildren. These projects do not try to understand the content of the chat between human users at a story level. Therefore, they differ from the goal of our research.

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.* 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 [29]. 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 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 [5]. AutoTutor is a tutoring system that helps students learn a subject through a conversation with an avatar with a talking head [11]. Spierling and Iurgel 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 [28]. 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 real users talk and create stories without constraints.

CONCLUSION

We have presented *Raconteur 2*, 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 story points, and developing a coherent story path with the support of relevant media files.

Future work will focus on modeling the storytelling dialogue, and better tailoring the story patterns to the user's intention. 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

1. BECKER, H., NAAMAN, M., AND GRAVANO, L. Learning Similarity Metrics for Event Identification in Social Media. In *Proc. of WSDM '10: the 3rd Intl. Conf. on*

Web Search and Data Mining, ACM Press, New York, New York, USA, 2010.

2. BLACK, J. B. AND WILENSKY, R. An Evaluation of Story Grammars. In *Cognitive Science*, vol. 3 (3), pp. 213-230, 1979.
3. BIRD, S. KLEIN, E, LOPER, E. AND BALDRIDGE, J. Multidisciplinary Instruction with the Natural Language Toolkit. In *Proc. of TeachCL '08: the 3rd Workshop on Issues in Teaching Computational Linguistics*, Columbus, Ohio, USA, 2008.
4. BURKE, M., MARLOW, C., AND LENTO, T. Social Network Activity and Social Well-being. In *Proc. of CHI 2010*, ACM Press, Atlanta, Georgia, USA, 2010.
5. CASSELL, J. Embodied Conversational Agents: Representation and Intelligence in User Interfaces. In *AI Magazine*, vol. 22 (4), pp. 67-84, 2001.
6. 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.
7. CHI, P.-Y. AND LIEBERMAN, H. Raconteur: From Intent to Stories. In *Proc. of IUI 2010*, ACM Press, Hong Kong, China, 2010.
8. CHURCHILL, E., GOODMAN, E., AND O'SULLIVAN, J. Mapchat: Conversing in Place. In *Ext. Abs. of CHI 2008*, ACM Press, Florence, Italy, 2008.
9. FELLBAUM, C. *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA, USA, 1998.
10. 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.
11. 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.
12. KIRK, D., SELLEN, A., ROTHER, C., AND WOOD, K. Understanding Photowork. In *Proceedings of CHI 2006*, ACM Press, Montréal, Québec, Canada, 2006.
13. KUIPERS, B. Commonsense Reasoning About Causality: Deriving Behavior from Structure. In *Artificial Intelligence*, vol. 24, pp. 169-203, 1984.
14. KUMAR, R, MAHDIAN, M., AND MCGLOHON, M. Dynamics of Conversations. In *Proc. of KDD '10: the 16th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, New York, NY, USA, 2010.
15. KURLANDER, D., SKELLY, T., AND SALESIN, D. Comic Chat. In *Proc. of SIGGRAPH 1996*, ACM Press, New York, NY, USA, pp. 225-236, 1996.
16. 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.
17. LABOV, W. Some Further Steps in Narrative Analysis. In *Journal of Narrative and Life History*, vol. 7 (1-4), pp. 395-415, 1997.
18. LINDE, C. *Life Stories: The Creation of Coherence*, Oxford University Press, 1993.
19. LIU, H. AND SINGH, P. ConceptNet: a Practical Commonsense Reasoning Toolkit. In *BT Technology Journal*, vol 22 (4), pp. 211-226, 2004.
20. LIU Y., SHAFTON, P. SHAMMA, D., AND YANG, J. Zync: the Design of Synchronized Video Sharing. In *Proc. of DUX '07: the 2007 Conf. on Designing for User Experiences*, ACM Press, Chicago, IL, 2007.
21. POLANYI, L. *Telling the American Story: A Structural and Cultural Analysis of Conversational Storytelling*, MIT Press, 1989.
22. 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 *Proc. of CHI 2010*, ACM Press, Atlanta, GA, USA, 2010.
23. SCHANK, R. C. *Explanation Patterns: Understanding Mechanically and Creatively*, Psychology Press, 1986.
24. SCHANK, R. C. *Tell Me a Story: A New Look at Real and Artificial Intelligence*, Northwestern University Press, 1991.
25. SHAMMA, D., SHAW, R., SHAFTON, P, AND LIU, Y. Watch What I Watch: Using Community Activity to Understand Content. In *Proc. of MIR '07: the Intl. Workshop on Multimedia Information Retrieval*, Ausburg, Bavaria, German, 2007.
26. SINGH, P., LIN, T., MUELLER, E. T., LIM, G., PERKINS, T. AND ZHU, W.-L. Open Mind Common Sense: Knowledge acquisition from the general public. In *Proc. of the 1st Intl. Conf. on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*, Irvine, CA, USA, 2002.
27. SPEER, R., HAVASI, C., AND LIEBERMAN, H. AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge. In *Proc. of AAAI2008*, AAAI Press, Chicago, IL, USA, 2008.
28. SPIERLING, U. AND IURGEL, I. "Just Talking about Art"- Creating Virtual Storytelling Experiences in Mixed Reality. In *Proc. of ICVS '03: the 2nd Intl. Conf. on Virtual Storytelling*, Toulouse, France, 2003.
29. 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.