NORTHWESTERN UNIVERSITY

Compelling Computation: Strategies for Mining the Interesting

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Computer Science

By

Sara Hodges Owsley

EVANSTON, ILLINOIS

December 2007

Acknowledgements

To everyone who has supported me throughout my academic career, thank you. Thank you Kris, Larry, and InfoLab members for your creativity, vision, and guidance. Thank you friends for keeping me entertained and happy during my graduate career. Thank you family for having confidence in me. Most of all, thank you Sanjay for your love and support.

Abstract

Compelling Computation: Strategies for Mining the Interesting

Sara Hodges Owsley

Living in a world where the machine and the Internet are ubiquitous, many people work and play online, in a world that is, ironically, often isolated and lonesome. While the Internet, as intended, connects us to information, products and services, it often draws us away from the rich connections that are created through interpersonal communication. The goal of this dissertation is to use the machine to connect people. Not to information, products or services, but to each other. I propose to use the machine, the very machine that pulls us apart, to bring us together, connecting people through stories. People tell stories as a way to be less lonesome, reaching out to people that they can relate to. The three systems that I describe in this dissertation are intended to facilitate and amplify connections between people through stories.

Buzz is a digital theater installation that autonomously finds and exposes the most emotional stories from the millions of blog postings on the Web, bringing the stories to life with a cast of virtual actors using speech synthesis. *Wrigley Buzz* is a variant of *Buzz* that seeks out stories about brands and products, connecting companies to consumers via stories about experiences with and opinions towards their products. Finally, *News at Seven* is an automatically generated news show that, in addition to giving the news, automatically finds and presents different points of view, connecting people through stories and opinions about current events expressed in the blogosphere. These three systems serve the goal using the machine to connect people, not to information or products, but to each other.

Table of Contents

ACKNOWLEDGEMENTS **3**

ABSTRACT **4**

CHAPTER 1: AN INTRODUCTION TO COMPELLING COMPUTATION **8**

CHAPTER 2: BUZZ **16**

COMPELLING STORIES 21
RETRIEVAL, FILTERING AND MODIFICATION MODEL 22
Query Formation *24*
Topics of Interest 24
Structural Cues 25
Blog Finding and Result Processing *27*
Candidate Extraction *27*
Story Modifiers *28*
Story Filters *28*
Relevance to Topics of Interest and Inclusion of Structural Story Cues 29
Complete Passages 29
Filtering Retrieval by Syntax 30
Content or Impact Modifiers *31*
Content or Impact Filters *32*
Filtering Retrieval by Affect 32
Colloquial Filtering 33
Language 34
Presentation Modifiers *35*
Presentation Filters *36*
Presentation Syntax Filter 36
Detecting Gender-Specific Stories 36
Evaluation *39*
CREATING A PERFORMANCE 40
The Display *41*
Director Level Control *41*
Compelling Speech *43*
CONCLUSION 45

CHAPTER 3: FINDING SENTIMENT IN TEXT **48**

DOMAIN SPECIFIC WORD CONNOTATIONS 49
RTS APPROACH 51
Training Data *52*
Domain Classifier *53*
Sentiment Query Formation *55*
Case Retrieval and Evaluation *56*
RTS System Evaluation *58*
TOPICAL SENTIMENT 60
Training *62*
Using Domain Knowledge *62*
Topic Word Windows *63*
Sentence Level Sentiment *63*
Pointed Verbs *64*
Levin Index 64
Stemming 65
Results and Coverage of Topical Sentiment Classification Approaches *65*
Combination Approach *66*
CONCLUSION 68

CHAPTER 4: TOPICAL STORIES, OPINIONS AND REACTIONS 69

TOPICAL STORIES 70
EMBODIED PRESENTATION OF STORIES 71
TOPICAL STORIES INVOLVING DRAMATIC SITUATIONS 75
OPINIONS 76
CONCLUSION 79

CHAPTER 5: COMPELLING NEWS STORIES AND PERSPECTIVES 80

FINDING OPINIONS RELATED TO NEWS STORIES 81
FORMING QUERIES 83
ASSESSING RELEVANCE OF CANDIDATE OPINIONS 84
ADDITIONAL FILTERS 86
OPINIONS ON NEWS STORIES 87
CELEBRITY NEWS 89
CONCLUSION 91

CHAPTER 6: NETWORK ARTS AND THE ASSOCIATION ENGINE 92

THE ASSOCIATION ENGINE 93
THE PATTERN GAME 94
Connected Thesauri 95
Word Familiarity 96
Context 99
Relation Types 100
The Pattern Game Generated by Digital Improvisers 100
ONE WORD STORIES 101
Generating Templates 102
Type Based Dictionaries 104
Selecting a Template 105
Selectional Restriction 106
The One Word Story Generated By Digital Improvisers 107
PRESENTATION 108
Interactive Model 109
A Team of Improvisers 110
Five Actors 111
INSTALLATIONS 115
CONCLUSION 116

CHAPTER 7: RELATED WORK 117

WHY STORY GENERATION? 118
EARLY STORY GENERATION SYSTEMS 118
RECENT APPROACHES TO STORY GENERATION 119
STORY DISCOVERY 120
RELATED ARTISTIC SYSTEMS 121

CHAPTER 8: CONCLUSIONS AND THE FUTURE 123

STORY DISCOVERY 125
STORY PRESENTATION 126
STORIES 127

APPENDIX 128

1. HANDLE WITH CARE: DIRECT MAIL AND THE AMERICAN DREAM BY GIRLCHARLIE 128
2. MSAPI VISEME TO IMAGE MAPPING 130
3. LEVIN VERB CLASS 31.2: ADMIRE VERBS (LEVIN 1993) 132

REFERENCES 134

Chapter 1: An Introduction to Compelling Computation

But a world without stories is fundamentally inhuman.

– Roald Hoffmann (Hoffmann 2000)

Imagine a world without stories. A world where history is lost and mistakes are repeated. A world where learning is restricted to theory without practice. A world where we are alone with our experiences, with no way to express our fears, gain compassion or empathize with the common experiences of others. A world in which nothing is imagined and creativity is stifled. A world where negative experiences are bottled up with no venue for catharsis or confession, and positive happenings go uncelebrated. A world where self expression lies in actions and facts. A world in which the chronicle of our life is forgotten. Such a world is not only hard to imagine, it is inhuman.

The importance of stories in our lives is immeasurable. Stories connect us. They are the structure of all human communication. They provide a medium through which we can form relationships with other human beings. These relationships and connections bring purpose and meaning to our lives. Without stories, we are lonesome.

> *We are lonesome animals. We spend all our life trying to be less lonesome. One of our ancient methods is to tell a story begging the listener to say -- and to feel -- Yes, that's the way it is, or at least that's the way I feel it. You're not as alone as you thought.*
>
> - John Steinbeck (Steinbeck 1954)

Playing such a variety of roles, stories take many different forms. Nonfiction stories are written to pass on knowledge, often peppered with legend and myth that evolves with the passing of time. News stories are reported to recreate the events of the day, making people aware of current issues facing others around the world. Fictional stories are told by parents to help develop creativity and fantasy in their children's minds. Beyond these types and many others, overwhelmingly, the most commonly told story is a first person experience.

It is first person stories that provide us with the most powerful way to make meaningful connections with the people around us. We can express our goals and fears, our experiences and struggles, our triumphs and happiness, and gain comfort, empathy, criticism, and support from others. We tell first person stories to put our thoughts out in the world and pass on knowledge and ideas, but most importantly, we seek emotional connections to other people. We want to know that others are having similar experiences, joys and frustrations in their lives and the world that we live in. It's not surprising that we listen to and read first person stories for reasons that are strikingly similar to why we tell them; for entertainment, to be touched, to gain comfort, to empathize, and to learn about and connect to the author.

> *To be a person is to have a story to tell.*
>
> - Isak Dinesen (Simmons 2002)

Our lives are filled with stories. We love to talk about what happens to us. The most straightforward reason that we talk about our experiences is that it is a way for us to remember things or to clarify for ourselves what has happened. Even when the listener or audience is inactive or unresponsive, simply telling a story can serve as catharsis and therapy. In an extreme case, Tom Hanks' character, in his 2000 movie *Cast Away*, gains comfort as he tells stories to an inanimate volleyball who he lovingly calls "Wilson."

> *Hey [to Wilson], you want to hear something funny? My dentist's name is James Spalding.*
>
> - Tom Hanks in *Cast Away* (Zemeckis 2000)

Living in a world where the machine and the Internet are ubiquitous, many people work and play online, in a world that is, ironically, often isolated and lonesome. While the Internet, as intended, connects us to information, products and services, it often draws us away from the rich connections that are created through interpersonal communication.

> *A literary journal is intended to connect writer with reader; the role of the editor is to mediate.*
>
> - John Barton

The goal of this dissertation is to use the machine to connect people. Not to information, products or services, but to each other. We propose to use the machine, the very machine that pulls us apart, to bring us together, connecting people through stories. Much like the editor of a literary journal, I propose a system that mediates a connection between people through stories.

What do we need in order to build such a system? Well, first we'll need a venue for people to publicly read and write stories. Luckily, such a venue already exists in the form of the weblog (*blog* for short). The popular rise of blogs came in 2001 as a venue for information, commentary, opinions, narrative, and stories, posted by everyday people in a journal format. Blogs are vastly published, with 70 million total blogs, and 15.5 million currently active, or updated in the past 90 days (BusinessWeek 2007; Technorati 2007). The widespread use of blogs is a testament to people's desire to tell their stories.

We also need a way for people to find the stories written in blogs that might be interesting to them, or might provide the connection that they are seeking. Current blog search engines provide users with a way to search for topical content. While these search engines span the entire blogosphere and return topically relevant blog entries, they fall short in scaffolding connections between people. First, they only return the most popular or authoritative blogs. Given the vastness of the blogosphere, if the most popular blogs are always returned, then millions will remain unseen or unread, unconnected. Also, blog search engines do just what their name implies, search blogs, and not all blogs take the form of stories. Current blog search engines on their own will not meet our goals of connecting people through stories.

> *Nobody wants to listen to what happened to you today unless you can make what happened appear interesting. The process of livening up an experience can involve simply telling that experience in such a way as to eliminate the dullest parts, or it also can involve "jacking up" the dull parts by playing with the facts.*
>
> - Roger Schank (Schank 1990)

Next, we need the system to find, not just stories, but interesting or compelling stories. Current blog search engines measure importance through popularity; these systems do not evaluate the emotional impact or interestingness of the blogs that they return. The system must be able to find stories that people will want to read or hear, not just the ones that are popular, but the ones that are emotional and compelling.

Finally, we need a way to present these stories to people; a way that compels them to listen and enables them to embed themselves in the narrative. While a textual format (book, magazines, blogs, etc) has historically been a common way to communicate stories, hearing a story told by a person is a much more powerful and expressive experience and allows the listener to truly empathize with and connect to the person telling the story. We must keep this in mind in the presentation of stories to people.

To address these issues we have built a system called *Buzz* that exists to enable emotional connections, connecting people with each other via stories. It reaches deep down into the blogosphere, beyond the popular and authoritative, and finds compelling stories; from real people who are telling stories and begging to be heard. These stories are emotional, touching, funny, surprising, comforting, eye-opening, etc. They expose people's fears, dreams, experiences, and opinions. The system changes the definition of relevance from similarity and popularity to emotional impact and interestingness. By connecting people to these stories, the system, *Buzz*, amplifies their voices. Being heard is not only beneficial for the blogger, but the viewer can gain comfort, have a laugh, relate to an experience, get advice, share in the blogger's joy or sorrow, take a walk in the their shoes, feel less lonely, hear an opinion or perspective, or simply be entertained. *Buzz* connects people.

Below (in Table 1-1) are three stories retrieved by the *Buzz* system. These stories were extracted from actual blog posts, using *Buzz's* retrieval, filtering and modification engine. While the three are all first person stories, it's clear that their authors had differing goals while writing the stories; the first and third, a "fight" and a "dream," are a therapeutic telling of their experiences, while the second, a "confession," was likely intended more as a self-descriptive narrative, exposing the blogger's political stance. While differing in content and goals, these three stories are all highly affective and compelling, and most importantly, they involve situations that everyone can relate to: fights, confessions, dreams, etc.

my husband and i got into a fight on saturday night; he was drinking and neglectful, and i was feeling tired and pregnant and needy. it's easy to understand how that combination could escalate, and it ended with hugs and sorries, but now i'm feeling fragile. like i need more love than i'm getting, like i want to be hugged tight for a few hours straight and right now, like i want a dozen roses for no reason, like a vulnerable little kid without a saftey blankie. fragile and little and i'm not eating the crusts on my sandwich because they're yucky. i want to pout and stomp until i get attention and somebody buys me a toy to make it all better. maybe i'm resentful that he hasn't gone out of his way to make it up to me, hasn't done little things to show me he really loves me, and so the bad feeling hasn't been wiped away. i shouldn't feel that way. it's stupid; i know he loves me and is devoted and etc. yet i just want a little something extra to make up for what i lost through our fighting. i just want a little extra love in my cup, since some of it drained.
I have a confession. It's getting harder and harder to blindly love the people who made George W Bush president. It's getting harder and harder to imagine a day when my heart won't ache for what has been lost and what we could have done to prevent it. It's getting harder and harder to accept excuses for why people I respect and in some cases dearly love are seriously and perhaps deliberately uninformed about what matters most to them in the long run.
I had a dream last night where I was standing on the beach, completely alone, probably around dusk, and I was holding a baby. I had it pulled close to my chest, and all I could feel was this completely overwhelming, consuming love for this child that I was holding, but I didn't seem to have any kind of intellectual attachment to it. I have no idea whose it was, and even in the dream, I don't think it was mine, but I wanted more than anything to just stand and hold this baby.

Table 1-1 A set of stories retrieved and presented by the *Buzz* System.

In addition to finding interesting stories, the *Buzz* system presents these stories in a novel way. Instead of showing the stories in their original form, as plain text, the system embodies the blogger with an avatar and generated voice (as depicted in Figure 1-1). These avatars present the stories by reading them aloud, with each actor attentive to the one currently reading. This enables a stronger connection with the viewer as it simulates the blogger telling you their story. Much like the difference between printed newspaper and television news, *Buzz* presents the stories in a different and more engaging manner.

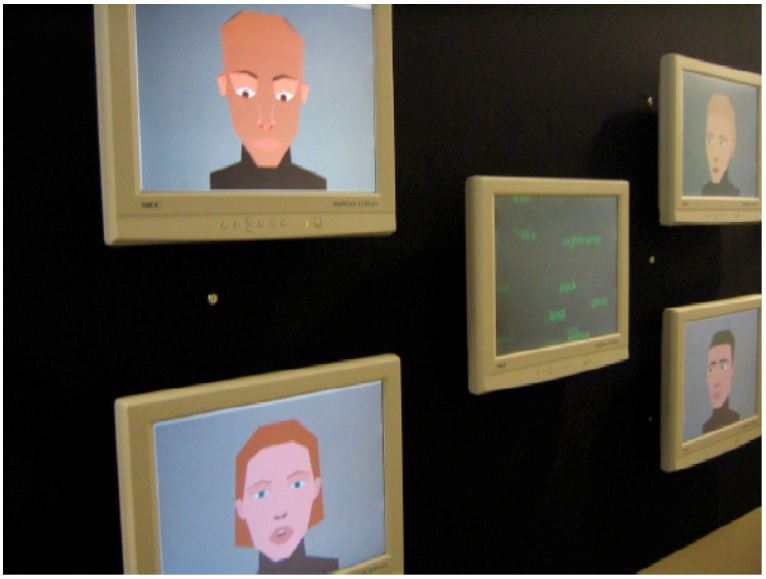


Figure 1-1 An installation of *Buzz* in the Ford Engineering Design Center at Northwestern University.

Now that I've established what needs to be done, this dissertation will take you on a journey through how we accomplish these tasks, how *Buzz* was built.

- In Chapter 2, I'll give an overview of *Buzz* as a digital theater installation, and what a user or viewer of the system experiences. I'll present the backend system that finds

interesting and compelling stories from the blogosphere, through a story retrieval, filter, and modification model.

- In Chapter 3, I'll explain our methods for identifying emotion and describe a sentiment classification system that we built to evaluate stories for the *Buzz* system.
- In Chapter 4, I'll present an additional use case for *Buzz*, finding emotional stories on topics, namely brands and products, and discuss the impact of this system in the marketing world. Through opinions and stories expressed in a natural manner in the blogosphere, *Buzz* is able to connect companies to their consumers and to the general public.
- In Chapter 5, I'll describe the use of *Buzz* as a method for finding perspectives and opinions on current events, as part of an automatically generated news show called *News at Seven*.
- In Chapter 6, I'll describe how *Buzz* fits into a larger area of work called *Network Arts* and was preceded by a system, called *The Association Engine*, which led to the creation of *Buzz*.
- In Chapter 7, I will situate *Buzz* and *Network Arts* in the space of related work as digital theater installations, story generation and story telling systems.
- Finally, in Chapter 8, I'll present what's in store for the future of *Buzz*.

Chapter 2: Buzz

A good story cannot be devised; it has to be distilled.

- Raymond Chandler

Buzz connects people via stories. It reaches deep down into the blogosphere, beyond the popular and authoritative, and finds stories; stories that are selected not by their popularity, but by their emotional impact. These stories are poignant, touching, funny, surprising, comforting, eye-opening, etc. They expose people's fears, dreams, experiences, and opinions. Instead of showing the stories as plain text, the system embodies the blogger with an avatar and generated voice, enabling a stronger connection with the viewer.

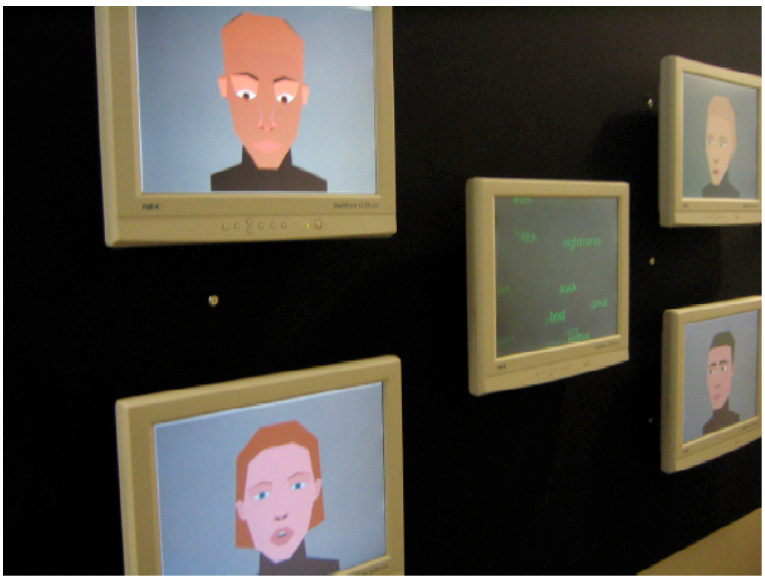


Figure 2-1 An installation of *Buzz* in the Ford Engineering and Design Center at Northwestern University.

Buzz (Figure 2-1) finds and presents these stories in the form of a digital theater installation. It discovers blogs which are compelling; those where someone is laying their feelings on the table, exposing a dream or a nightmare that they had, making a confession or apology to a close friend, or regretting an argument that they had with their mother or spouse. It embodies the author (blogger) with virtual actors who give voice to these stories. The focal point of the installation displays the most emotional and evocative words from the monologue, shown as falling text.

To understand what it is like to experience *Buzz*, consider a viewer who approaches it in a public installation. As he draws near the installation he sees five monitors on a wall in the shape of an X (Figure 2-1). He notices four faces occupying the outer monitors; and hears one of these characters speaking, appearing to tell a story to the others. The character is speaking about a dream she had last night (see Table 2-1 for the full dream), and the others are closely listening to this story, turned and facing her. As she speaks, her words fall on the central screen (Figure 2-2) – love, life, dream, and friend – highlighting the most moving words in her story. As her story concludes, the character on the screen above her takes over, lightheartedly confessing about his poor singing skills (see Table 2-1 for the full confession). The performance goes on in this manner indefinitely as the characters continue to tell compelling stories.

So, I had this dream last night of someone who used to be in my life. And I really, truly loved him. And he's been away now for more than twice as long as I even knew him. But I still miss him. I still love him. And I'm fairly sure there is a part of him that still loves me. He was my best friend.
I have a confession to make. I can't sing, I can't dance, I can't do nothing. cause I had no professional training, neither do I have the talent to begin with. I can't hit the high notes and worst of all, I don't know how to sing any song. McDave and Famezgay, on the other hand, being karaoke veterans, sung up lots of songs with ease, while I had to wail and screech to keep up with the melody. So sorry for making you guys suffer from my preposterous vocals. It was very unfortunate that I love to sing but I can't sing.
Keith and I got into a fight last night when I got home from Kentucky…. he said I was just like my father…. which honestly is the worst insult any human being can bestow upon me…. I tried to roll over and just go to sleep, but it hurt, I won't lie, it really hurt to have to hear those words…. "I'm just like my father"……OUCH. Before long though, back up the stairs he came, said he was sorry for saying that and then he told me he was going to sleep on the couch…but it occurred to him that this would be the last time he'd have the opportunity to lay beside me in OUR bed. It was to be the last time we'd sleep side by side. That statement was very final.

Table 2-1 Three sample stories found and presented by the *Buzz* system.

Figure 2-2 A close-up of the central screen of a *Buzz* installation.

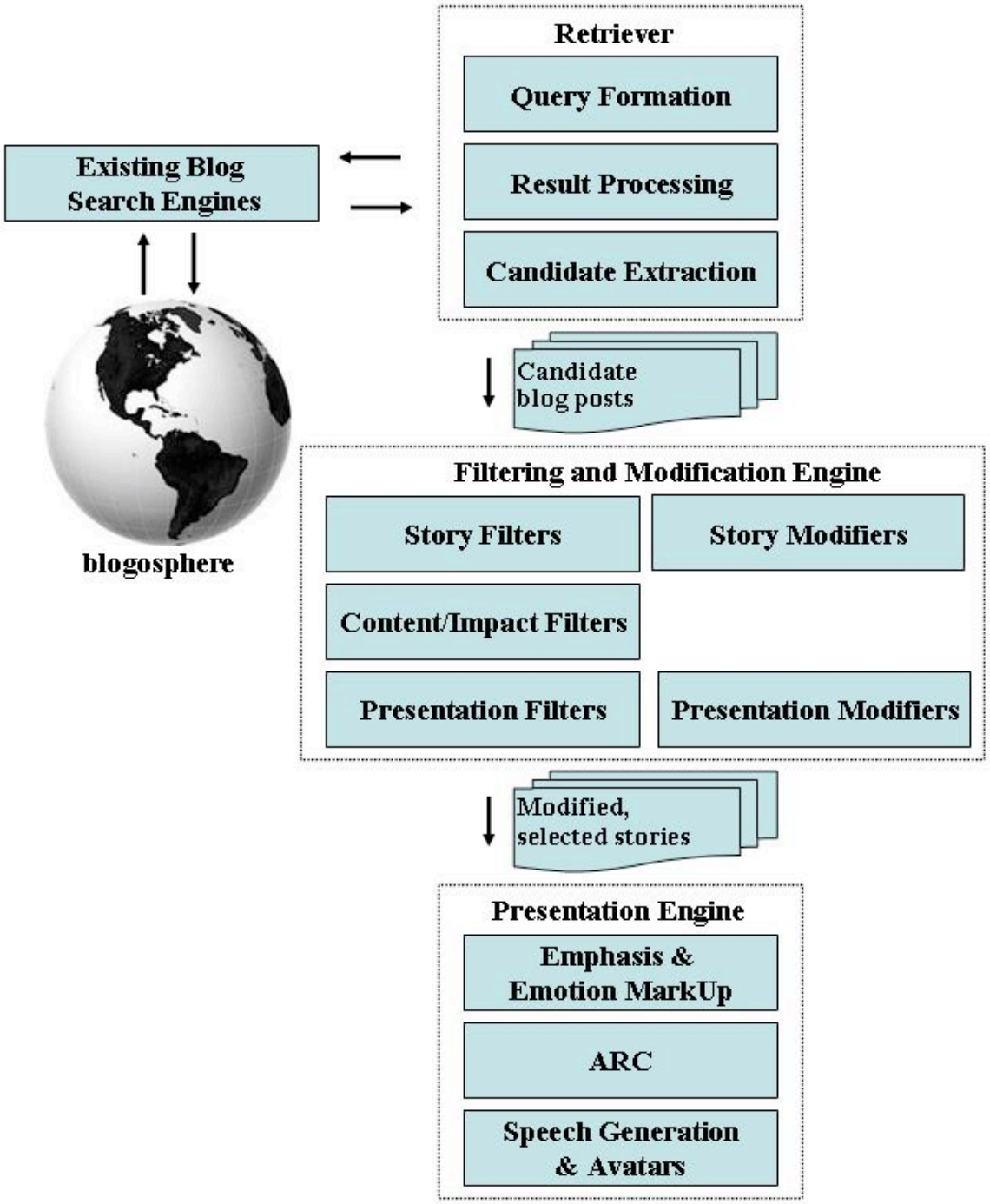


Figure 2-3 An architecture diagram of the *Buzz* system.

Figure 2-3 illustrates how *Buzz* works from the formation of queries to the final presentation of stories. To find compelling stories, the system mines the blogosphere (the global corpus of blogs), collecting posts where the author describes a dramatic and compelling situation: a dream, a nightmare, a fight, an apology, a confession, etc. After retrieving these pages, the system extracts candidate stories from the entries. The system then takes the stories through a set of modifiers, aimed at transforming the candidates to make them look more like stories, be more emotionally amplified, and sound appropriate when spoken by animated characters. After being transformed, the candidate stories are passed through a set of filters, aimed at focusing the system on candidates that take the syntactic form of a story, are emotional amplified, and will sound appropriate when presented by an avatar; filtering out candidates that do not meet these criteria. Overall, the filtering engine is highly selective, discarding 98% of the retrieved candidate stories. These techniques are critical to the final performance as they ensure that the stories found will engage the audience.

After passing through the filtering and modification engine, the resulting story selections are emotion-laden and compelling. Next, the stories go through a markup stage, to prepare them for performance by an animated character. Several techniques are used to give the presentation of the stories a realistic feel and to make performances engaging to an audience. The story is marked up for speech and animation cues in a number of ways. It is marked up at a sentence level by a mood classifier, providing cues to the avatar and generated voice as to the affective state of the story as it progresses. This markup also includes emphasis and timing cues to yield better cadence and prosody from computer-generated voices.

Finally, the performance is planned and guided by dramatic Adaptive Retrieval Charts (or ARCs) are used to provide a higher level control of the performance, similar to that of a director.

These ARCs allow for various performance types from the most basic – a single virtual actor telling an individual story, for example as part of an online system – to more complex – for example an ongoing performance of multiple virtual actors in a physical installation. These charts specify performance needs at the level of story type, topic, gender, length, etc and are used to drive the retrieval of the stories. While these charts drive the performance, the virtual actors themselves are also driven by instructions, in a sense, on "how to act." For example, they are attentive to the actor currently speaking by turning and looking at them, and they pause in the appropriate places to make the performance feel more realistic.

COMPELLING STORIES

A first pass at building *Buzz* revealed that the content of blogs is incredibly wide-ranging, but unfortunately often very dull. *Buzz* succeeded in finding stories that were on point to any provided topic, but the results were not compelling. People blog about a wide range of topics; for example, their class schedule, what they are eating for lunch, how to install a wireless router, what they wore today, and a list of their 45 favorite ice cream flavors. While this is interesting to observe from a sociological point of view, it does not make for a compelling performance. Not only are the blogs on these topics boring, but the length of the blog posts vary widely from one sentence to pages upon pages, and most do not take the form of a story or narrative.

We need to give the system strategies for finding stories that will be compelling and engaging to an audience. To do so, the system employs a model for the aesthetic qualities of a compelling story. These qualities include but are not limited to:

1. on an interesting topic
2. emotionally charged

3. complete and of an appropriate length to hold the audience's attention

4. familiar to an audience

5. involving dramatic situations

6. comprised of developed characters

We designed *Buzz* to find stories with all of these qualities.

RETRIEVAL, FILTERING AND MODIFICATION MODEL

In building *Buzz*, I developed a model for the retrieval, filtering, and modification of stories that takes advantage of the vast size of the blogosphere, aggressively filtering the retrieval of stories. First, the system retrieves a large set of blogs using existing blog search engines. The retrieval process includes a query formation stage, retrieval of blog using existing search engines, result processing, and the extraction of candidate blog posts. Following this stage, the candidate blog posts are sent to the filtering and modification engine. In this engine, candidate stories are extracted from the posts and filtered using many different metrics. The stories that pass through all these filters are known to be impactful and appropriate stories for presentation in a *Buzz* performance.

There are three functional categories for *Buzz's* filtering strategies:

- *Story filters* are those which narrow the blogosphere down to those blog posts that include stories, including strategies that make use of punctuation, relevance to topics, inclusion of phrasal story cues, and completeness to indicate a text that is likely to have a dramatic point. For example, the *List filter* removes candidate stories that contain more than 3 commas per sentence as they are perceived as taking the form of a list as opposed to a story.

- *Content* or *impact filters* are used to find interesting and appropriate stories - those with elevated emotion, familiar, and relevant content that is free of profanity and other unwanted language use. For example, the *Affect filter* runs each candidate story through an emotional classifier and discards those stories that are classified as emotionally neutral. This filter enables the system to present emotional stories.
- *Presentation filters* are used to focus on content that will sound appropriate when spoken through a computer generated voice, and presented by an avatar of the appropriate gender. For example, the *URL filter* discards candidate stories that contain URLs as they sound awkward when presented by a computer generated voice.

In addition to filters, there is also a set of modifiers that alter the text of the retrieved stories.

- *Story modifiers* alter the text so that the structure looks more like a story. For example, the *Structural Story Cue modifier* truncates the story such that the structural story cue appears in the first sentence of the story; these cues are intended as natural starting points for stories.
- *Presentation modifiers* change the text to make it sound more appropriate in spoken as opposed to written form. For example, the *Abbreviation modifier* transforms abbreviations to the expanded form to ensure that they sound more appropriate when spoken by the avatar.

All filters and modifiers are configurable to adjust to different installations or deployments. The integration of this model into the overall *Buzz* system can be seen in Figure 2-3. In this diagram, notice that the modifiers and filters exist in a single module, illustrating the integrated nature in which these pieces interact. The following sections describe the use of these filters, modifiers and retrieval strategies in the overall system.

Before diving into the details of this architecture, I'd like to address why we chose to handle this problem the way we did, as there are many other approaches we could have taken. Given the vastness of the blogosphere, our system typically has on the order of hundreds or thousands of candidate stories for any one topic. This allows us to use simple techniques (predominantly search technology), building simple filters to be highly selective among these candidates. While these filters will have low recall in retaining good candidates, the precision is quite high in that the stories that are retained are quite good. We could have approached this problem using more complex techniques, but the vastness of the corpus of stories made that unnecessary.

Query Formation

Two types of query formation strategies are used in the *Buzz* system; one strategy that uses popular topics found daily on the web, and another uses a library of structural story cues to seek texts that take the form of a story. While both are described below, we have found the latter to be significantly more effective in yielding stories that are of interest to a large audience.

Topics of Interest

A compelling story is generally about a compelling topic, one that interests the audience. For this reason, we chose the day's most popular searches from Yahoo, provided by Yahoo Buzz (Yahoo! 2006), as topics. Search engines recently began to provide a log of their most frequently used query topics. Yahoo! takes this a step further to provide the most frequently queried topics in a set of categories. Their categories currently include: Overall, Actors, Movies, Music, Sports, TV, and Video Games. In the Actors category, the top three topics from March 7,

2007 are “April Scott,” “Lindsay Lohan,” and “Jessica Alba.” In the Overall category, the top three topics from March 7, 2007 are “Britney Spears,” “Antonella Barba,” and “Anna Nicole Smith.” These feeds worked well as a seed to story discovery, as we were using the topics that people were searching for most and discovering people’s thoughts and opinions on these topics. The intuition here is that the topics that people search for most are likely the topics that they are most interested in.

We found Wikipedia (Wikipedia 2005) to be another source for topics of interest as the site maintains a list of “controversial topics”. The list shows topics that are in “edit wars” on Wikipedia as contributors are unable to agree on the subject matter. This list includes topics such as “apartheid,” “overpopulation,” “ozone depletion,” and “censorship.” These topics, by their nature, are topics that people are passionate about. One March 7, 2007, Wikipedias “List of controversial issues” included such topics as “Bill O’Reilly,” “Abortion,” “Osama bin Laden,” “Stem Cell Research,” “Censorship,” “Polygamy,” and “MySpace.”

Using these two sites as sources for topics, these topics are used as queries and sent to a set of existing blog search engines. Using topics of interest as the source of topic keywords and blogs as the target, we are able to discover what was being said today about what people were most interested in today.

Structural Cues

Through experiencing *Buzz* in the world and watching audiences’ reactions and responses to stories, we discovered more generalized traits of compelling stories. The most compelling stories to watch or hear were those in which someone is laying his or her feelings on the table, exposing

a dream or a nightmare that they had, making a confession or apology to a close friend, or regretting an argument that they had with their mother or spouse.

Codifying these qualities, we built our story discovery engine to seek out these types of stories. We added a component to the retrieval that forms queries based on a structural story cue. These cues are designed to find instances in which a writer is starting to tell a story in the form of a dream, nightmare, fight, apology, confession, or any other emotionally fraught situation. Such cues include phrases such as "I had a dream last night," "I must confess," "I had a terrible fight," "I feel awful," "I'm so happy that," and "I'm so sorry," etc. The most straightforward structural story cue would be if the author said, "I have a story to tell you," or even "Once upon a time."

This realization was an important turning point in our system's capabilities with regard to retrieving compelling stories. The newest instance of *Buzz* no longer focuses on the popular or contentious topics, but instead focuses on stories in different types of emotion laden situations (dreams, fights, confessions, etc.).

These stories are more interesting as the blogger isn't merely talking about a popular product on the market, or ranting about a movie; they are relaying a personal experience from their life, which typically makes them more emotionally charged. The experiences they describe are often frightening, funny, touching, or surprising. They describe situations which have a common element in all of our lives, allowing the audience to embed themselves in the narrative and truly connect with the writer, whereas the topically based approach excluded the portion of the audience that was not familiar with the topic at hand (a popular actress, story in the news, etc.).

In the 19th century, a French Writer, Georges Polti enumerated 36 situational categories into which all stories or dramas will fall. These included modern categories such as vengeance, pursuit, abduction, murderous adultery, mistaken jealousy, and loss of loved ones (Polti and Ray

1940). While the language describing these situations now sounds somewhat dated, the concepts behind these situational categories bear a resemblance to the types of stories that might be interesting to hear.

Including structural story cues as a search parameter not only gets us to more interesting story topics and content, but we also tend to see more character depth and development in the stories. As writers describe dramatic situations in their lives, more pieces of their personality and personal issues with themselves and others around them are revealed as a result.

Blog Finding and Result Processing

The queries formed in the previous stage, such as "I had a dream last night," are sent to a set of existing blog search engines. The system collects the top *n* results (where n is configurable and is currently 1000) from these engines. Each result contains a title, summary, and URL of a blog related to the given query. The system filters duplicate results and non-blog results (i.e., user profile pages). Next, the HTML content for each blog result is retrieved.

Candidate Extraction

The content for each blog result may contain multiple posts which may or may not be relevant to the query. In order to generate a set of candidate stories, the system must be capable of extracting the individual posts that are relevant to the query. To identify the relevant posts within the blog result, all "text" tags in the HTML of the blog entry are removed, that is, tags used to alter the look of text such as the italics tags, the bold tag, the underline tag, and the anchor tag. After removing these tags, the system finds all occurrences of the given query terms and structural story cues on the page. For each occurrence, it searches for the last previous

occurrence of and the next occurrence of a natural breaking point. The section between these two points is taken as a candidate story. The natural breaking points before and after a piece of text will often be tags that divide paragraphs, so the algorithm will accomplish the goal of finding the relevant paragraphs. In combination with the filters described in the following sections, this algorithm works quite well as a generalized method for extracting candidate stories from online sources.

Following the candidate extraction step, what remains is a set of candidate stories, ready to be sent through the filtering and modification engine.

Story Modifiers

Story modifiers are modification strategies aimed at transforming the candidate story into a more story-like structure. The main strategy in this category involves the structural story cues described in the query formation section. While these cues are initially used as a method to retrieve stories, they are also used to truncate the blog post into the section that structurally is most like a story. Often blog posts are retrieved that include the structural story cue, but it occurs in the middle of a paragraph. Since the stories are initially divided by paragraphs in this system, story cues would not actually occur at the beginning of the candidate story. To change this, this modifier truncates the story to begin with the sentence that includes the structural story cue. The end result are stories that take the form laid out in the structural story template, beginning with phrases such as “I had a dream last night,” or “I got into a fight with.”

Story Filters

Story filters are those which narrow the blogosphere down to those blog posts that include stories, including strategies that make use of punctuation, relevance to topics, inclusion of phrasal story cues, and completeness.

Relevance to Topics of Interest and Inclusion of Structural Story Cues

This filter is intended to evaluate the relevance of candidate stories to the topics of interest and/or the structural story cue used in their retrieval, filtering irrelevant candidates. In the case of a topic of interest query, the candidate stories are phrasally analyzed, eliminating posts that do not contain at least one of the two word phrases (non-stopwords) from the topic. For example, given a topic of 'Star Wars: Revenge of the Sith,' entries that contained the phrase 'star wars' were acceptable, but not entries that merely had the word 'star' or 'wars.' The remaining blog entries were thought to be relevant to the current popular topic. In the case of a structural story cue query, the candidate story is analyzed to ensure that the story queue is present and occurs in the first sentence of the story. This ensures that the structural cue is used as intended, to start the story.

Complete Passages

In finding stories, the system must ensure that it finds complete stories, that is, ones that outline a complete thought. Finding stories that are complete passages involves finding complete thoughts or stories of a length that can keep the audience engaged. For the most part, we found that blog authors format their entries in a way such that each paragraph contains one distinct thought. Under this assumption, the paragraph where the structural story cue and/or topic are mentioned with the greatest frequency will suffice as a complete story for our system. Given the

method described to extract candidate stories from blogs, these candidate stories will likely take the form of a complete paragraph. If this paragraph is of an ideal length (between a minimum and maximum character and word threshold), determined by viewing *Buzz* with stories at many different lengths, then it is proposed as a candidate story. For a public installation of *Buzz*, we found that stories between 150 and 600 characters long were ideal. Again, given the large volume of blogs on the web, letting many blogs fall through the cracks because they are too long or too short is acceptable for our purposes.

Filtering Retrieval by Syntax

In our first pass at retrieving stories from blogs, we noticed that the system often found lists or surveys instead of text in paragraph form. For example, one blogger posted an exhaustive list of lip balm flavors. Others posted answers to a survey about themselves (their favorite vacation spot, favorite color, favorite band and actor, etc.). These are clearly not good candidates for stories to be presented in a performance.

To solve this problem, we chose to filter the retrieved candidate stories by syntax. Stories that met any of the following indicators were removed as they often signify a list:

1. too many newline characters (currently more than six in a four hundred character block)
2. too many commas (currently more than 3 in a sentence or more than 1 in 15 characters)
3. too many numbers (currently more than 1 number - no longer than 4 continuous digits - in a sentence)

This method successfully filters blog entries that contain a list or survey of some sort. While the recall of stories that pass through this filter based on syntax is lower than other methods, the system is optimized for precision so that we are confident that the remaining stories do not

contain lists or surveys. Given the large volume of blogs on the web updated every minute, letting some potentially good blogs fall through the cracks sufficed for the system's purposes.

Content or Impact Modifiers

Content or Impact Modifiers are modification strategies aimed at transforming the candidate story to make it more compelling or impactful. The current strategy in this category involves removing extraneous content in order to focus on the spines of the stories. For example, if the story contains any parenthetical, bracketed or braced content, it is removed. This includes any remaining html or xml tags. This is based on the notion that if you were reading this post to a friend, you might ignore such content as it breaks up the flow of the story.

In addition to the above mentioned *Content/Impact modifiers*, I envision another modifier in this category, called an *amplifier.* An *amplifier* would alter the candidate stories so that they are more impactful, emotional or colloquial. This system would transform words that occurred in a story to more emotional words with the same connotation. The end result would be a story that conveyed the same meaning, yet with more emotional impact than in its original form.

This could be implemented with a combination of a part of speech tagger, a connected thesaurus and our Naïve Bayes sentiment classification model (discussed in the next chapter). The system would attempt to replace adjectives in the candidate story, namely ones that have only one sense in the connected thesaurus, making the word unambiguous. From the synonym set, it could chose the synonym with a higher "sentiment magnitude" from the sentiment classification Naïve Bayes model. This "sentiment magnitude" is a calculation of how emotion-bearing a term is. This system will scale and be configurable for how much to amplify a story.

Content or Impact Filters

Content or *Impact filters* are used to find interesting and appropriate stories; those with elevated emotion, familiar, and relevant content that, if desired, is free of profanity and other unwanted language use.

Filtering Retrieval by Affect

Given that our initial version of *Buzz* was reading blogs that were boring or uninteresting, and since such a large volume of blogs exist on the web, we strove to filter the retrieved relevant stories by affect, giving us the ability to portray the strongest emotional stories. We accomplish this using a sentiment classification system. Sentiment analysis is a modern text classification area in which systems are trained to judge the sentiment (defined in a variety of ways) of any document. Since we simply want to know if a story is emotional or not, we define sentiment as valence, that is, how positive or negative a selection of text is.

In our sentiment classification system, a combination of case-based reasoning, machine learning, and information retrieval approaches are used (Owsley, Sood et al. 2006; Sood, Owsley et al. 2007). Using a training set of labeled data (movie and product reviews), our system is able to judge how emotional words are based on their appearance in the training data, and then use the most emotional words in a story as a query to a case base of labeled data. While many others have used such data to build Naïve Bayes sentiment classifiers, we find that using a case based approach preserves the large differences in affective connotation of words across domains. For example, while the word "cold" has a very negative connotation in describing a person, being "cold" is seen as a positive attribute of a beverage.

We collected a case base of 106,000 movie and product reviews labeled with a star rating between one and five (one being negative and five being positive). We omitted reviews with a score of three as those were seen as neutral. We built a Naïve Bayes statistical representation of these reviews, separating them into two groups, positive (four or five stars) and negative (one or two stars). Given a target document, the system creates an "affect query" as a representation of the document. The query is created by selecting the words with the greatest statistical variance between positive and negative documents in the Naïve Bayes model. The system uses this query to retrieve "affectively similar" documents from the case base. The labels from the retrieved documents are used to derive an affect score between -2 and 2 for the target document. This tool was found to be 73.39% accurate.

For *Buzz*, blogs which score between -1 and 1 are seen as neutral and not good candidates for a performance. When using the emotional filtering tool, *Buzz* is considerably more compelling. The actors are able to retrieve stories from the Web based on emotional stance, enabling the theatrical agents to juxtapose positive and negative stories on the same topic. This tool will be discussed in depth in the next chapter.

Future work on this classification tool includes creating a model of affect based on Ekman's six emotion model (*happiness*, *sadness*, *anger*, *disgust*, *fear*, *surprise*) (Ortony, Clore et al. 1987; Ekman 2003; Liu, Lieberman et al. 2003). This would allow for greater control of the flow of the performance arc through emotional states.

Colloquial Filtering

Shamma, et al. (Shamma, Owsley et al. 2004), began exploring the use of Csikszentmihalyi Flow State (Csikszentmihalyi 1991) as a method of keeping the audience engaged through

audiovisual interaction. In *Buzz*, for an audience to stay engaged, they must understand the content of the stories that they are hearing. That is, the story can't involve topics that the audience is unfamiliar with or contain jargon particular to some field. The story must be colloquial. The story must also not be too familiar as they audience could get bored or lose interest.

To determine how familiar a story is, I built a classifier that makes use of page frequencies on the web. For each word in the story, the system looks at the number of pages in which this word appears on the web, a frequency that is obtained through a simple web search. Applying Zipf's Law (Zipf 1949), the system can determine how colloquial each word is (Shamma, Owsley et al. 2004). A story is then classified to be as colloquial as the language used in it. Given a set of possible stories, colloquial thresholds (high and low) are generated dynamically based on the distribution of scores. If more than n percent of the words in a story fall below the minimum threshold (where n is configurable, currently n is 5), that story is seen as being too obscure and is discarded.

Language

It's important that that language used in a candidate story is appropriate for presentation through *Buzz.* For example, it would sound awkward for a *Buzz* digital actor to present a story that begins "In my last post" as the story is taken out of the context of the blog; or for an actor in a public installation to read a story that contains profanity. For this reason, *Buzz* uses a language filter that can be configured to remove stories which include profanity, or even stories which include words that expose the fact that it was extracted from a blog. For example, some blog posts are often started with the phrase "In my last post…" While this is appropriate when a

reader understands that what they are reading is a blog, etc., this is inappropriate or awkward when taken out of the context of the blog posting, and presented through an embodied avatar.

To filter out stories with such language, this filter uses a dictionary based approach. It can be provided with a list of words to filter based on. From there, the system can be configured to only filter based on those words, or to also include stems of those terms for broader coverage. As with all other *Buzz* filters, this filter may be turned "on" or "off" when appropriate.

Presentation Modifiers

In addition to *Presentation filters*, I built a set of *Presentation modifiers* aimed at altering the text to make it more appropriate for presentation through a computer generated voice. Upon reaching the *Presentation modifiers*, the candidate stories have passed through the three major filter sets (story filters, content or impact filters, and presentation filters) as well as the story modifiers. The next step is to prepare them to be spoken by a speech generation engine.

Adjacent punctuation is condensed as the speech engines use this punctuation for pauses, so adjacent punctuation would result in long pauses. Any remaining numbers, dates, and monetary amounts are altered to be readable by the speech engines. Finally, abbreviations are substituted to their expanded form, and any remaining acronyms or abbreviations are expanded to instruct the speech engine correctly. For example, "APA" would be expanded to "A. P. A." so that the speech engine spells out the acronym as opposed to treating it as a word.

Upon completing all of these modifications, the story candidate is passed through all filters a second time. This ensures that any transformations made on the text did not change its value or quality as a story, or how appropriate it is for presentation.

Presentation Filters

Presentation filters are used to focus on content that will sound appropriate when spoken through a computer generated voice, and presented by an avatar of the appropriate gender.

Presentation Syntax Filter

While syntax filtering was included in the "Story Filters," it is also important in the Presentation Filters, due to the limitations of computer generated speech. Blogs, by their nature, are often casually punctuated and structured. While this isn't generally a problem for the reader, it poses a problem when presented through a text-to-speech engine. Text-to-speech engines use punctuation as cues for prosody and cadence. (Sproat, Ostendorf et al. 1998) For this reason, when a story is poorly punctuated, or it contains too many numbers, numbers with many digits, URLs, links, or email addresses which sound bad when presented by a text-to-speech engine, they are filtered by the presentation syntax filter.

This filter also removes stories that contain a direct quote which makes up more than one third of the story. We found that lengthy direct quotes are awkward when read by a computer generated voice. When a person reads a direct quote, they often have a change of inflection to indicate a different speaker. This change does not occur in computer generated voices, often causing the listener some confusion. For this reason, candidate stories that fall into this category are discarded.

Detecting Gender-Specific Stories

One problem encountered in a first pass of building *Buzz* was that gender-specific stories were occasionally read by actors of the incorrect gender. For example, if a blogger describes

their experiences during pregnancy, it is awkward to have this story performed by a male actor. Conversely, if a blogger talks about their day at work as a steward, having this read by a female could also be slightly distracting.

As a solution to this problem, I sought to detect and classify gender-specific stories. Unlike previous gender classification systems (Koppel, Argamon et al. 2003), it was not necessary for our system to classify all stories as either male or female. Rather, it was only important for the system to detect stories where the author's gender is evident, thus classifying stories as male, female, neutral (in the case where gender-specificity is not evident in the passage), or ambiguous (in the case where both male and female indicators are present).

To do this, the system looks for specific indicators that the story is written by a male or a female. These indicators include self-referential roles (roles in a family and job titles), physical states, and relationships. These three types of indicators are treated as three separate rules for gender detection in the system.

To detect self-referential roles in a blog, the system looks for 'I' references including "I am", "I was", "I'm", "being", and "as a." These phrases indicate gender-specificity if they are followed within five words (if none of these five words are pronouns) by a female-only or male-only role such as wife, mother, groom, aunt, waitress, mailman, sister, etc. Such roles were collected from various sources and enumerated as such. This rule set is meant to detect cases such as "I am a waitress," which would indicate that the speaker is a female. Excluding extra pronouns between the self reference and the role eliminates false positives such as "I was close to his girlfriend," where the additional 'his' ensures that this rule is not applied.

To detect physical states that carry gender connotations, the system again looks for 'I' references, as above, followed within five words by a gender-specific physical state such as

"pregnant." This rule is meant to detect cases such as "I am pregnant." As in detecting roles, we also ignore cases with extraneous pronouns between the 'I' reference and the physical state. This eliminates false positives such as "I was amazed by her pregnancy."

To detect male or female-only relationships, the system looks for use of the word 'my' followed within five words by a male or female only relationship such as husband, ex-girlfriend, etc. This rule is intended to catch cases such as "my ex-husband." Again, cases with extraneous pronouns are ignored to eliminate false positives such as "my feelings towards his girlfriend." In this our first pass at a gender specific story classification system, we make the assumption of heterosexual relationships, which we hope to relax in a future system.

If any of three above indicators exists in a story, and they agree on a male/female classification, then the story is classified as such. If they disagree, it is classified as 'ambiguous.' If no indicators exist, it is classified as 'neutral.'

This gender detection tool was evaluated using a corpus of 96 stories retrieved by *Buzz*. These stories were retrieved from an indexed corpus of stories found by *Buzz*. They were selected by queries for words that often indicate gender-specificity ('pregnant', 'mom', 'mother', 'dad', 'father', 'girlfriend', 'boyfriend', 'husband', 'wife', and 'daughter'). They were manually sorted into three groups, stories written by females, males, or neutral (written by males or females). This sorting was based on textual cues that gave a clear indication of gender, and was verified unanimously by from five participants.

While this gender-classification system is still simple, it does an admirable job. Results showed that the gender detection tool performed very well, as seen in the precision and recall scores in Table 2-2. Overall precision and recall were both approximately 91.67%. Enabling

Buzz with the ability to detect and handle gender specific stories has created a more realistic performance, without the distraction of an actor performing a gender-mismatched story.

Document Type	**Precision**	**Recall**
Female-specific	92.59%	86.21%
Male-specific	100%	84.62%
Gender-neutral	89.66%	96.30%
Overall	91.67%	91.67%

Table 2-2 Precision and Recall Scores for detection of gender specific stories.

Evaluation

An example of three stories discovered by *Buzz* can be seen in Table 2-1. The stories shown were retrieved and passed through all above mentioned filters. To evaluate the effectiveness of *Buzz's* filters in finding compelling stories, I conducted a user study including twelve participants. Each participant was given five stories to score on a scale from one to ten (uninteresting to interesting). The stories were chosen at random from a set of stories selected by *Buzz* as good candidates for a performance, and a set of stories retrieved by *Buzz* but removed as they did not pass one of the filters.

On a scale of one to ten (uninteresting to interesting), the study participants found *Buzz* selected stories to be an average 7.13 and *Buzz* rejected stories to be an average of 4.3. A graph of the frequencies of participant scores across *Buzz* accepted and *Buzz* rejected stories can be seen in Figure 2-4.

While I am quite pleased with the results of the *Buzz* story discovery system as a method for autonomously discovering compelling stories, I think it has even broader implications and contributions. Consider how people tell stories – they never "create" or "invent" new stories, instead they recall the most interesting stories they've heard or experiences they've had, they may merge different experiences together, and they alter the details to make a more compelling story. In a sense, we've codified how people tell or create stories, building a system to do just that. In a way, I think this system has made contributions to the decades old problem of story generation in Artificial Intelligence.

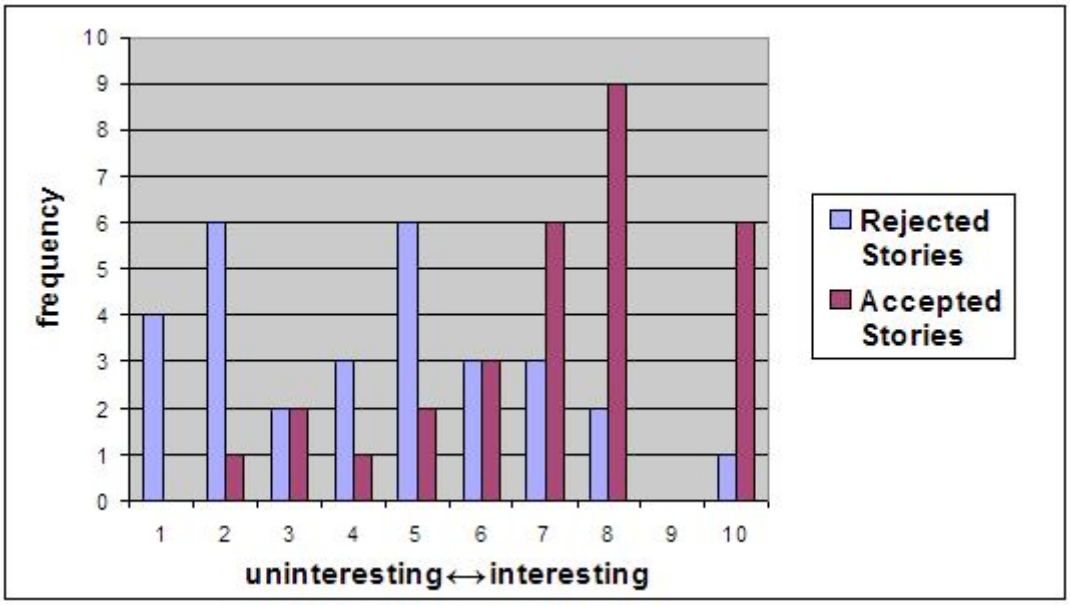


Figure 2-4 The results of an evaluation of the *Buzz* story discovery engine, where participants judged stories on a scale from 1 to 10 (uninteresting to interesting).

CREATING A PERFORMANCE

While finding compelling stories is an important aspect of *Buzz*, conveying them to an audience in an engaging way is just as crucial. I found several aspects of the presentation to be critical. The performance must follow a dramatic arc that keeps the audience engaged. Text-to-

speech technology and graphics must be believable (or suitable) and evocative. While these issues are a subset of those critical to an engaging performance, I chose to address these directly as I feel that our findings can generalize to other performance systems.

The Display

The current *Buzz* installations include five flat panel monitors in the shape of an 'x'. The four outer monitors display actors represented by different adaptations of the graphics from Ken Perlin's Responsive Face technology (Perlin 1996; Perlin and Goldberg 1996). These faces (Figure 6-4 and Figure 6-5) are synchronized with voice generation technology (NeoSpeech 2006) controlled through the Microsoft Speech API, matching mouth positions on the faces to viseme events, lip position cues output by the MSAPI (see Appendix 2 for more details on the visemes used). Within this configuration, the actors are able to read stories and turn to face the actor currently speaking.

The central screen (Figure 2-2) displays emotionally evocative words, pulled from the text currently being spoken, falling in constant motion. These words are extracted from the stories using the emotion classification technology described in the section on "Filtering Retrieval by Affect" and in the following chapter. The most emotional words are extracted by finding the words with the largest disparity between positive and negative probabilities in a Naïve Bayes statistical model of valence labeled reviews. I've found this display to be a good addition to the actors to give the audience additional context in the performance and amplify the impact of the emotional words.

Director Level Control

Given the above classifiers and filters, the system is able to retrieve a set of compelling stories. These filters and classifiers also give us a level of control of the performance similar to that of a director. Having information about each story such as its "emotional point of view", its "familiarity," and the likely gender of its author, the structure of an ongoing performance or individual story presentation in an online system can be planned out from a high level view before retrieving the performance content, giving the performance a flow, based not only on content, but on emotion, familiarity, on-point vs. tangential, etc. Given a topic, when the system is presenting multiple stories, the system can juxtapose stories with different emotional stances, different levels of familiarity, and on-point vs. off-point. These affordances give a meaningful structure to the performance.

To provide a high level control of the performance, we created an architecture for driving the retrieval of performance content. The structures, called Adaptive Retrieval Charts (or ARCs), provide high level instructions to the *Buzz* engine as to what is needed, where to find it, how to find it, how to evaluate it, how to modify queries if needed and how to adapt the results to fit the current goal set. To get an idea of how the ARCs interact with the blog search and filters, see Figure 2-5.

The pictured ARC (Figure 2-5) defines a point/counterpoint/dream interaction between agents. The three modules define three different information needs, as well as the sources for retrieval to fulfill these needs. The first module specifies that we want a blog entry that is on point to a specified topic, has passed through the syntax and colloquial filters, and is generally happy on the topic. The module specifies using Google Blog Search (Google 1996) as a source. The source node specifies to form queries by single words as well as phrases related to the topic.

If too few results are returned from this source, we have specified that queries are to be continually modified by lexical expansion and stemming.

The ARC extensible framework allows for interactions from directors with little knowledge of the underlying system. In a future system, we will accomplish this via a range of possible interfaces from storyboarding and affect manipulation to a natural language interface.

```
- <arcs>
  - <arc name="point/counterpoint/dream">
    - <module name="opinion 1" type="blog entry" length="less than 60 words" filter="syntax + on point +
        colloquial + happy">
        <source name="Google Blog Search" location="www.blogsearch.google.com"
          queryFormationStrategy="single term + phrasal" evaulationStrategy="popularity"
          queryModificationStrategy="lexical expansion + stemming" adaptResults="text only" />
      </module>
    - <module name="opinion 2" type="blog entry" length="less than 60 words" filter="syntax + on point +
        colloquial + angry">
        <source name="Google Blog Search" location="www.blogsearch.google.com"
          queryFormationStrategy="single term + phrasal" evaulationStrategy="popularity"
          queryModificationStrategy="lexical expansion + stemming" adaptResults="text only" />
      </module>
    - <module name="dream" type="blog entry" length="less than 80 words" filter="syntax + funny + off
        point + colloquial + dream">
```

Figure 2-5 A sample dramatic ARC used to drive a *Buzz* performance.

Compelling Speech

While text-to-speech systems have made great strides in improving believability of generated speech, these systems are not perfect (Black 2002). Their focus has been on telephony systems, where the length of time of spoken speech is limited and emotional speech is unnecessary. In watching a performance of *Buzz* using such text-to-speech systems, the voices tend to drone monotonously during stories longer than one to two sentences. An additional problem is caused by the stream of consciousness nature of some blogs, resulting in casual formatting with poor or limited punctuation. As mentioned earlier, text-to-speech systems generally rely on punctuation

to provide natural pauses in the speech. In blogs where limited punctuation was present, the voices tended to drone on even more.

In response to these issues, we created a model for emotional speech emphasis. Others have created models for how to emphasize words in generated speech (Raux and Black 2003) and which words to emphasize. While these models are successful, we strove to create a simple model that would scale to our needs and capture the emotional element of the stories. Our system also includes a model for emotional speech emphasis at the sentence level. First, the system uses a sentence level emotion classifier to determine which sentences in a story are highly affective, and which emotion they are characterized by. In the exemplary installation of *Buzz*, the text is marked up at the sentence level for its emotional content (happy, sad, angry, neutral, etc.). This can also be done in larger spans such as at the paragraph or story level, or in smaller spans such as the word or phrase level. The models of emotion used can be replaced by a more or less detailed model of emotion.

Many speech engines allow XML markup to control the volume, rate and pitch of the voices, as well as to insert pauses of different periods (specified in milliseconds) in the speech. We use this XML markup, in combination with an off-the-shelf audio processing toolkit, to alter the sound of the speech according to its emotional markup. For example, to handle a happy sentence, the pitch will be raised, rate will be increased, and the pitch of the voice will rise slightly at the end of the sentence. These changes in speech attributes to portray emotion are informed by a study of emotional human speech (Cahn 1990). For each of Ekman's six emotions, Cahn's study provides dimensions on which generated speech can be altered in order to render that emotion.

An off the shelf audio processing toolkit was necessary for this system as some speech engines to not support the necessary parameter of control (pitch, rate, volume, etc.). In addition, using an audio tool as post processing often conserves the quality of the voice more than using the speech engine's internal controls.

In addition to using a model of emotional emphasis, the system inserts pauses into the audio stream at natural breaking points. This technique tends to improve performance on blogs with limited punctuation. While our model of emotional speech emphasis is simplistic, we've found it to be an effective way to enhance the *Buzz* experience. We expect to further tweak our emphasis model in response to audience or user feedback.

CONCLUSION

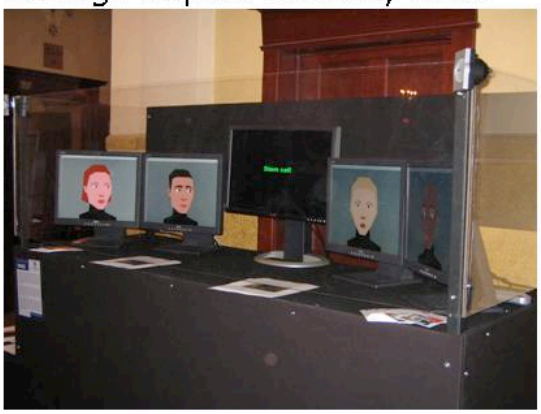
Initially, story discovery within *Buzz* was based on popular topics. As I approached the task of engaging the user, it became more important that the stories themselves were compelling, as opposed to being topical. Using filters, modifiers and information retrieval strategies that focused on finding the interesting and not the topical has resulted in an engaging theatrical installation.

While finding compelling stories to present is a very important part of the *Buzz* performance, presenting these stories in a way that is meaningful and engaging is equally important. I found issues of gender-specificity, voice prosody, and presentation flow and order to be the aspects of a *Buzz* performance with which I could make great strides in improving. Future work in the presentation of *Buzz* will include more realistic looking avatars and continued work on enhancing the voice prosody.

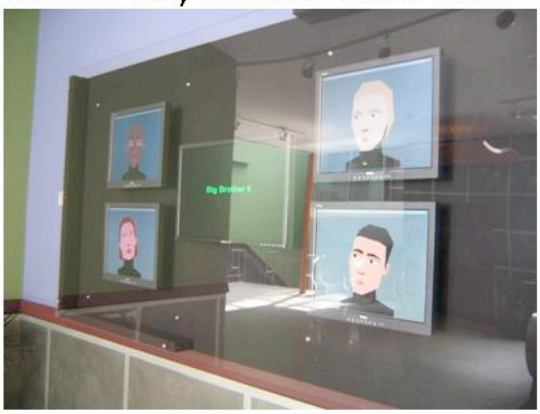
Enabling *Buzz* with the ability to discover compelling stories has produced great results. *Buzz* has changed from a research project accessing stories that were unbearably dull, exposing the boring nature of many blogs, to a system that engages its viewers. The performance is now not driven simply by the relevance of on-line content, but by the blogger's emotional state. The highly emotional content engages the audience and creates a high visibility installation.

Buzz debuted from April 22nd through May 1st 2005 at the Athenaeum Theater as a part of the 8th Annual Chicago Improv Festival. It was well-received by actors, writers, producers and theater-goers alike during this ten day installation. *Buzz* was installed in the lobby of Chicago's Second City Theater at 1616 N. Wells St. in Chicago on August 24th, 2005 for a one year installation. *Buzz* was also exhibited at Wired NextFest in New York City from September 29th to October 1st, 2006. See Figure 2-6 for still shots of the four major installations.

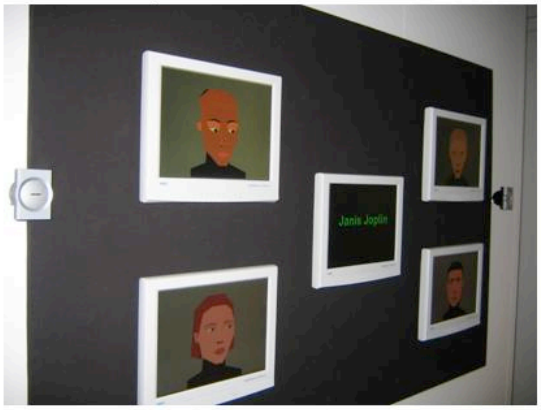
Chicago Improv Festival, 2005



Second City Theater 2005-2006



Ford EDC, Northwestern Univ.



Wired Magazine's NextFest, 2006

Figure 2-6 Pictures of four public Buzz installations.

Chapter 3: Finding Sentiment in Text

> *You want to write smart stories? Cool. You will automatically appeal to the (unfortunately) smaller herd that reads for smart/literary/quality/complex allusion kinda stuff. You want them to find a larger audience? You want them to be read widely? You want access to the genre herd? The herd wants emotion. The herd is a bunch of sappy, horny, schmaltzy, bored, angry, frustrated, romantic, thrill-seeking, guys and girls who will pick up a book, any book, that pushes their particular genre button well. They'll happily tolerate dumb, but I'm pretty sure they'll also tolerate smart if it comes along with a skillful push of their emotional button.*
>
> (RomanceBlogReader 2007)

Buzz only serves its function of connecting people via stories, if the stories that it presents are able to engage the viewers. Viewers will not become engaged in the presentation of narratives about what a blogger ate for breakfast, or directions for how to install a wireless router; rather, they can connect with bloggers who describe life experience such as marriage, dreams, and death, stories filled with emotional expression. To be successful in connecting people, *Buzz* must have the ability to distinguish between emotional and non-emotional stories. To make this judgment, we enabled *Buzz* with the ability to evaluate the sentiment of the stories that it retrieves.

Sentiment classification is a modern text classification problem that has been the focus of many researchers in recent years. The end goal of some sentiment classification systems would be to classify text into categories characterized by a model of emotion, such as Ekman's six emotion model (*surprise*, *anger*, *fear*, *happiness*, *disgust*, *sadness*). Within the scope of this project, we consider sentiment as valence, that is, how positive or negative a piece of text is. For *Buzz*, knowing if a story is very positive, very negative, or neutral is quite useful in terms of understanding its emotional impact, and in the end, a good metric for how interesting or compelling telling that story could be when performed.

DOMAIN SPECIFIC WORD CONNOTATIONS

Critical to the success of sentiment classification systems is addressing the notion that words are ambiguous, with different meanings in different domains or contexts. Words used to describe things in a positive light in one domain are likely to be a distinct and sometimes opposing set of words than in another domain. When describing a car in a positive way, you might use words such as "sleek" and "maneuverable"; but when describing a vacation destination, you might call it "relaxing" or "adventurous." These words are examples of positive words specific to domains. This means that there are different words that are important to conveying sentiment in different domains. Now consider a word such as "cold," where the emotional connotation is actually opposing across domains. A "cold politician" has a negative connotation, but describing a beverage as "cold" is usually positive. So, not only are words ambiguous, but some words have differing emotional connotations across domains (Finn and Kushmerick 2003; Owsley, Sood et al. 2006).

Given this discrepancy, in order to appropriately classify the sentiment of text from different domains using a machine learning approach, domain specific sentiment classifiers are necessary as the difference in meaning of words between domains reduces the accuracy of classification. A domain-specific approach, however, requires training data in every domain to be encountered in classification. Aue and Gamon surveyed various techniques for sentiment classification in new domains and concluded that labeled training examples from the new domain are needed for accurate classification (Aue and Gamon 2005). Enumerating and finding the appropriate labeled data in every single domain is time consuming and not scalable to all the domains one my encounter on the Web. There is also the problem of determining the correct classifier to apply for text where the domain is not known beforehand.

Approaching text classification from a different angle, systems have been built that use case-based reasoning to classify documents such as e-mail spam – a task often approached using machine learning techniques. Cunningham, et al, argue that case-based classification works well when there is a variation among individual cases that cannot be captured in a general statistical model (Bruninghaus and Ashley 1998; Cunningham, Nowlan et al. 2003; Healy, Delany et al. 2005).

While many researchers have built large scale statistical sentiment classifiers based on training data, we propose a case based component that preserves the differences in affective connotations of words across domains. We feel that a case-based approach to sentiment classification yields better accuracy by leveraging individual cases as opposed to a unified statistical representation. We believe it's necessary to use domain specific language to classify the emotional content. In combination with this case based approach, we use a statistical representation of the case base in order to determine which words are emotionally salient, as a

basis for case retrieval. This approach outperforms previous work and we feel that it accurately captures the domain specificity of language.

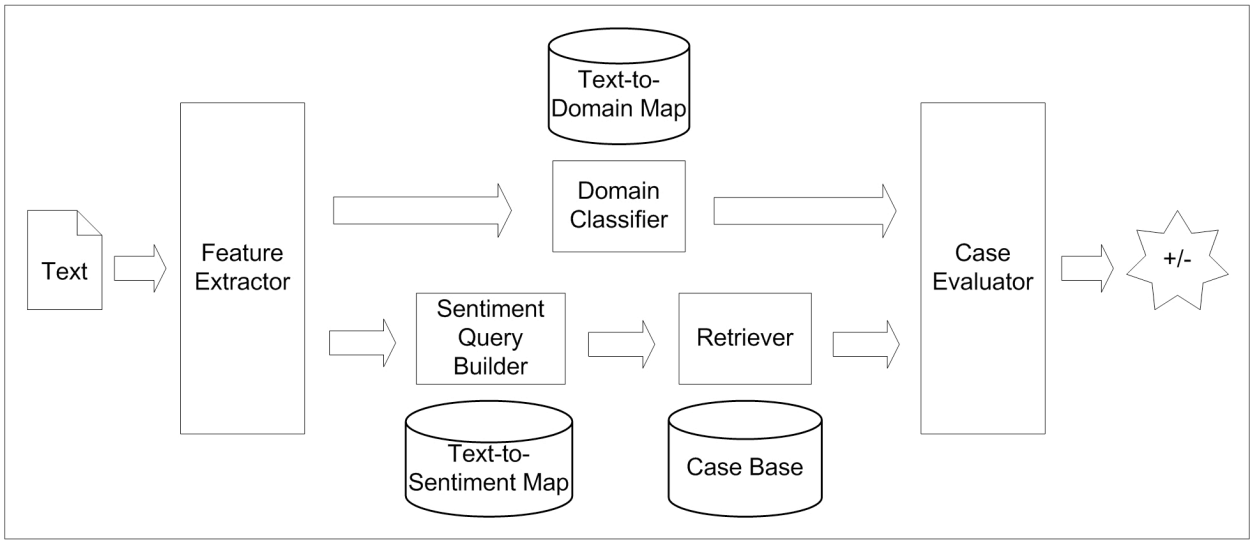


Figure 3-1 An architecture diagram of the RTS emotional classifier.

RTS APPROACH

Our sentiment classification system uses a combination of machine learning, information retrieval techniques, and case base reasoning to determine the sentiment of a piece of text. The name of the system, *RTS*, stands for *Reasoning Through Search* which makes reference to the search component of the system. As mentioned previously, we consider sentiment as valence, that is how positive or negative a piece of text is. Input to the system is a piece of text, and the system's output is a score between -2 and 2 that represents the valence of the text.

Figure 3-1 shows an architecture diagram of the RTS system. The overall system has four major steps. Given a document to classify, it first forms a "sentiment query" from the document, comprised of the most emotion bearing terms. The query is formed in this manner because we are searching for "emotionally similar" cases, not the standard "topically similar" cases. In

parallel, it classifies the domain of the document, returning a ranked list of likely domains. Next, it posts the sentiment query to a case base which retrieves emotionally similar labeled cases (movie and product reviews with known valence as number of stars). Finally, it uses the retrieved results to calculate a score by combining the ranked list of likely domains with the set of labeled cases to calculate a weighted average valence, the final result.

Training Data

The machine learning and case base components of our system required a set of labeled data for training. We gathered data from Rate-It-All (RateItAll 2006), an online repository of consumer written reviews on a wide variety of topics including: products, services, famous people, sports teams or figures, movies, music, and colleges. The reviews each have an associated rating, one to five stars, assigned by the author. Once submitted to RateItAll, the reviews do not go through an editorial process and are presented as is. We chose a subset of domains from RateItAll that we felt covered a breadth of topics. The domains we selected were: actors, books, colleges, destinations, drinks, electronics, food, movies, music, restaurants, software, sports and video games.

We collected a total of 106,961 reviews from these 14 domains. Some reviews consisted of a star rating with no review text or a very short review text, so we limited our collection to reviews with six or more words. The average length of a review was 47.86 words, with a minimum length of six words and a maximum length of 1205 words. We chose to build the data set with reviews in either extreme (positive and negative) in order to get a concentration of emotional language. Given that the reviews were rated between one and five stars, we labeled the set of

negative reviews as those with one or two stars, and positive reviews were those with 4 or 5 stars; reviews with 3 stars were ignored as they were viewed as neutral.

While this corpus of reviews is very useful as a training corpus, it does have some anomalies. Since the reviews do not go through any editorial process, they often contain misspellings, use slang words, and are off topic. Reviewers occasionally make mistakes in terms of the number of stars they assign to a review. Since RateItAll exists as a social network as well, the reviews often contain dialog between reviewers. As with any free text, human-generated content, such noise is unavoidable.

This corpus of reviews is used in all three major aspects of the system. First it is used as a corpus from which to extract emotion bearing terms to be used in the query formation stage. Next, it is used as training data for a Naïve Bayes Domain Classifier. Finally, it functions as a case base of labeled reviews.

Domain Classifier

We've established that words have different affective connotations across domains. Since our case base contains reviews across all fourteen domains, retrieving emotionally similar cases requires knowledge of the domain(s) that the current document is associated with. To meet this need, we built a Naïve Bayes domain classifier. Given a document from an unknown domain, the classifier returns a list of domains, ranked and weighted from most related to least related, where the weights are normalized probabilities.

The training data for this classifier was the set of reviews described in the previous section. To train the classifier, we treated each word as a feature of a document, using Porter's stemmer (Porter 1980) to unite words with morphological variation. Other machine learning researchers

have shown that use of other features (adjectives, phrases, etc.) provide an improvement in performance in similar systems (Muresan 2001; Bekkerman and Allan 2004), however, we found the performance of a system using words as features to perform well enough.

Each training document, *d*, was split into a vector containing the *n* unique words that appeared in the document, $d_i =< w_1, w_2, w_3, ... , w_n >$. The *frequency* with which each word appeared in a single document was ignored as we were building the system based on the mere *presence* of the words in the documents, allowing us to count the number of documents that each term appears in. Using the known domain of the document in combination with its word vector, we were able to generate the word probability tables for each domain. For each domain, this table is comprised of a word and the probability that that word appears in a document in that domain.

Given a document that the system is to classify, a ranked and weighted list of domains is created by calculating the NB probability that the document *d* is a part of each candidate class *c*, where *c* is one of the domains. The standard NB classifier uses a product of probabilities; however, to prevent underflow of this calculation, we used the summation of the logs of the probabilities. We also used add one smoothing to prevent the length of the document or the amount of training data in each domain from influencing the classification. The following equation was used to get the NB probability that the document *d* is a part of each class *c*:

$$P_{NB}(c \mid d) = P(c) + (\sum_{i=1}^{m} \log(P(w_i \mid c)^{n_i(d)}))$$

Equation 3-1

In a previous section I established that words have differing emotional connotations in different domains. If this were solely true, then it would imply that a document couldn't be classified accurately unless there was training data from that domain. In practice, we've found that some domains are related, using similar language to describe things in a positive and negative light. For example, words that are used to describe a movie in a positive light such as "gripping," "scary" and "funny" could also be used to describe a book in a positive way. This allows us to use a selection of domains that we think has broad coverage, as opposed to having training data from every possible domain. The domain classifier allows the system to figure out the closest likely domains for a document, even if the true domain of the document isn't represented in the training data.

Sentiment Query Formation

To find similar labeled reviews in our case base, we first have to turn our document into a query. In this case, we don't want the usual topical or content similar cases, but instead, we want cases with similar sentiment. In order to achieve this, we transform the document into a query of its most sentiment bearing terms. These terms have the largest difference in probabilities of occurring in positive documents versus negative documents in our review corpus.

To do this, we tokenize and stem the words in the document we are transforming. For each word *w*, we calculate its *sentiment magnitude* as follows:

$$sm(w) = \sum_{i=1}^{m} abs(\log(P_i(w \mid pos)) - \log(P_i(w \mid neg)))$$

Equation 3-2

where *i* denotes a domain in which we are examining the probabilities, with *m* domains in total.

Words with high *sentiment magnitude* are seen as being discriminating terms for emotion.

After calculating a *sentiment magnitude* for each word in the review, the system forms a query by sorting the words in descending order of their *sentiment magnitude*. We use the top four words as candidate words for the query. From experimentation with our case base, we found that four word queries performed best as a balance between retrieving too few and too many results. In many cases, when used in context, the sentiment bearing terms have opposing meaning. To account for this, for each candidate word, the system looks at the three words that occur before that word in the document. If a modifier (very, not, too, etc) is found within this window, the word is expanded to the phrase that includes the modifier. For example, the term 'funny' with a high *sentiment magnitude* would be transformed to 'not funny' if it is used that way in the document the system is trying to classify.

The generated four term query is then used to retrieve sentiment related cases from our case base. In some rare cases a generated query does not return results because the representation created for the document does not match any cases in the case base. In these instances, the query is relaxed and resubmitted to the case base by removing terms from the end of the query. Removal of words creates a more general query that has a higher likelihood of matching other cases.

Case Retrieval and Evaluation

In order to retrieve emotionally similar cases from the case base, the system must include an index and retrieval system. Instead of implementing our own retrieval system, an off-the-shelf search engine was sufficient for our purposes. We used Apache Lucene (Apache 2007), an open-

source search engine, to index and retrieve cases. All labeled reviews described above in the "training data" section were indexed in the engine.

The standard Lucene setup was used, indexing both the content of the reviews, as well as the known domain of the review. This allowed us to retrieve cases based on textual similarity to the review and to filter them based on the domain of the review. The Lucene engine was originally setup to ignore stopwords during indexing. This meant that searching for the phrase "too funny" would yield no results, though the phrase did exist in the case base, because the word 'too' was not included in the engine's index. We modified the indexing process so that all words, including stopwords, were indexed, allowing us to include phrases, such as those described in the previous section, in our queries.

When retrieving reviews the system needs to be able to quickly access the domain and star rating of the review. We strategically named the review files such that it contained meta-data including its domain, the rating assigned by its author (one to five stars), and a unique identifier (e.g. music- 4-8273.txt) allowing the system to access this information directly. The body of the text file contained the text of the review.

After a set of sentiment similar cases have been returned from the case base, a valence score is calculated based on the scores (star ratings) attached to the retrieved cases. We only utilize the top 25 results returned from Lucene when calculating a valence score as those are found as most "emotionally similar" to the target document. The sentiment classification of the document using related cases that are from a similar domain are likely to be more accurate. For this reason, the ranked and weighted domain list generated in domain classification is used to weight each sentiment score for the returned cases.

The overall valence score for the document is calculated as a weighted average of the scores from the retrieved cases. Given that *ls(c)* returns the labeled score of a case *c* and *w(c)* returns the weight of the labeled domain of a case *c*:

$$sc(d) = \frac{\sum_{i=1}^{m} (ls(case_i) - 3) * w(case_i)}{m}$$

Equation 3-3

calculates the sentiment score of a document *d* based on *m* retrieved cases (where *m* was typically set to 25). As previously mentioned, resulting valence score ranges from -2 to 2, instead of the 1 to 5 scale on the training reviews (accomplished by subtracting three in the equation above).

RTS System Evaluation

To evaluate this approach, we used a central set of training and testing data in order to implement other various approaches for performance comparison. We also ran a small human study to get a sense of how good humans are at the task of classifying the sentiment of a review.

The human study involved a questionnaire containing a random selection of 20 reviews from the RateItAll data set (excluding reviews with a score of 3, viewed as neutral). The 13 participants were asked to say whether each review was positive or negative. In scoring the questionnaires, the number of stars assigned by the review author was used as truth data, where reviews with 4 or 5 stars are positive and 1 or 2 stars are negative. After dropping the worst and

the best participant scores, the study showed that the average human accuracy at this task was 78.6% (15 out of 20).

We implemented two other common approaches to sentiment classification as points of comparison. The first is an "all-data" Naïve Bayes approach. This approach does not use domain knowledge, but is simply a Naïve Bayes sentiment classifier that is trained on all data. To test this approach the training data from a testing review's domain was excluded from the classifier for each test. Given that we have inherently have training data for all of the domains which our testing data falls into, this removal of in-domain training data was done in order to simulate the idea that using this approach, you won't have training data from all domains.

The other approach was an "ensemble" approach, comprised of a sentiment classifier built for each domain. The idea of an ensemble is to allow each classifier to vote on the final score, and in doing so, avoiding the problems that arise with large "all-data" classifiers such as the average out of disparate cases and data.

To test the *RTS* approach, we used the same training and testing data used in testing the systems above. We also excluded the in-domain training data to give a fair comparison to other systems. The results can be seen in Table 3-1 below. Our *RTS* approach outperformed the *Ensemble* and *All Data* approaches, while falling slightly short of the human baseline. We hope that future improvements to the system can close the gap to the human baseline performance.

Approach	Accuracy
Human Baseline	78.60%
RTS	**73.39%**
Ensemble	60.66%
All Data	66%

Table 3-1 A comparison of the performance of the RTS System against humans and two other automated approaches.

TOPICAL SENTIMENT

For *Buzz*, knowing the sentiment at the level of an entire story sufficed. However, when looking for emotional opinions on products, brands, or any topic, classification at the story or paragraph level was not appropriate (see Chapter 4). The general sentiment of a paragraph could be positive, yet around a specific topic in that paragraph might be neutral (as the topic was mentioned in passing) or even quite negative. For this reason, I began to investigate topical emotion in text. To explore this area, I implemented and evaluated a few various approaches, changing and testing the classification features. For training and testing corpora, I used the training data of reviews described previously, but pruned it down to only those reviews that contained the literal topic name (a movie, an actor, a politician, etc.) within the review as we can only evaluate the sentiment around a topic when a literal topic name is mentioned.

In considering the sentiment of a review, topical sentiment should more directly correspond to the number of stars given by the reviewer. That is, a review could have generally positive language in it, but the sentiment about the product at hand could be negative and marked with 1 star by the author, or vice versa. As an example of this, the following is a review (see Table 3-2)

from RateItAll on a song called “Just You N' Me” by the band *Chicago*. The author of this review, gave the song “one star.”

“I remember when I first heard Chicago (then called Chicago Transit Authority, as I recall) I thought they were brilliant. Looking back, that's a little strong even for their very early work which is very good. By the time they got around to recording this piece of cowflop, however, they had lost all claim to being real musicians.” - review by irishgit (http://www.rateitall.com/i-954786-chicago-just-you-n-me.aspx)

Table 3-2 An example of a generally positive review, which is negative about the focus topic.

The general language in this review is positive and sentimental about the band *Chicago*. However, toward the end, the reviewer states their dislike for the song being reviewed. His negative opinion about this song, which is the topic at hand, guided his decision to give the review a one star rating. For this reason, our corpus of reviews lends itself better as training and truth data for topical sentiment classification. Using a topic specific approach, the system might be more able to associate the label (number of stars) to the appropriate part of a review.

In the following sections, I will describe various methods for topical sentiment classification that I attempted and at the end will show results and discuss which methods were most accurate and effective. All of these approaches are grounded in the *RTS* method that I described in the previous sections and are focused on narrowing down the text used in classification to those terms that will be most accurate for sentiment on a topic.

Training

Given that we are now focusing on the sentiment surrounding a topical word or phrase in a piece of text, I thought it appropriate to test the effects of training the *RTS* system from the full corpus of reviews (including those that do not include the topic name) versus training on the pruned corpus of reviews that contain the literal topic name while only using the sentence(s) where the topic occurred for training. My hypothesis was that different emotional language may be used at the topical sentence level versus throughout the review and this distinction could be captured by limiting the training data.

Using Domain Knowledge

In the first part of this chapter and in previous work, I established that domain knowledge can be very beneficial to the accuracy of a sentiment classifier as words often have different connotations across domains. To test this notion for topical sentiment, I tested each approach both using the known domain knowledge of the review and not using the known domain knowledge to classify the testing data.

Using the known domain knowledge included three major modifications to the classification algorithm. First, the sentiment query is formed using the positive and negative word probability tables for the known domain, as opposed to using this probability calculated with data from all domains. Functionally, this means that in determining which words are "sentiment bearing" terms in the target document, we judge them based on which terms are sentiment bearing in the specific domain, as opposed to generally (across all domains). Next, when the query is submitted to the Lucene engine, it has an additional query term added which restricts the domain of the results ("domain:cars"). When no results exist in the specified domain, the Lucene engine

relaxes that constraint. Finally, if there are results returned from the Lucene engine in the domain specified, those results alone are used to calculate the final valence score, as opposed to using the ranked domain list result from the domain classifier. This is based on the intuition that truth data (the known domain) should be used before the domain classifier results.

Topic Word Windows

Intuition might tell you that the words immediately before and immediately after a topical mention in a piece of text might more accurately portray the writer's sentiment towards that topic. This could include verbs or adjectives describing the topic or the writer's feelings towards that topic. To capture this, one simple approach is to limit the sentiment classification to words within an *n* word window surrounding the topic. To complete this, all *n* word windows surrounding occurrences of the topic in the text to be classified are combined to form a document. This document is then scored using the standard *RTS* approach.

Sentence Level Sentiment

Based on this same intuition, another approach is to consider only words which occur in the sentence(s) where the topic occurs. All sentence(s) where the topic occurs are combined to form a document which is then scored using the standard *RTS* approach. Again, this approach assumes that the words within the sentence with the topic are more telling of the author's opinion on this topic, and that all other sentences only add "noise" to the classification.

Taking this notion a step further, another method is to only evaluate words of particular part(s) of speech that occur in the sentence where the topic occurs. This is under the assumption that there are syntactic qualities of words which determine their emotional impact or any

meaning for that matter. Many researchers have asserted that this is true (Levin 1993) and the most common assumption is that adjectives carry the most emotional weight (Kamps, Marx et al. 2004). Others have added that verbs and or adverbs carry significant value as well (Chesley, Vincent et al. 2006; Benamara, Cesarano et al. 2007). I'll address various combinations of parts of speech to test how well each characterizes the author's sentiment towards the topic.

Pointed Verbs

A first pass at these methods confirmed that adjectives at the sentence level were most effective for classifying topical sentiment. However, we felt that there are some verbs are very powerful in conveying ones feelings towards an object. Verb such as "like," "dislike," "love," and "hate" should not be ignored as they seem to very clearly express emotional opinions toward an object. For this reason, we wanted to have a hybrid method of classification which made use of not only adjectives, but pointed verbs, that is, verbs that were semantically tied to the idea of giving an opinion.

Levin Index

To generate such a corpus of pointed verbs, we look to the Levin Verb Index (Levin 1993), a listing of 3,401 verbs classified into both semantic and syntactic classes. We experimented with using various semantic and syntactic classes of verbs to form the set of pointed verbs used for classification. The set we found to be most indicative of emotional meaning is the verb class "31.2 Admire Verbs" which consists of 45 unique verbs (shown in Appendix 3) such as "like," "enjoy," "abhor," and "hate."

Because there were so many different combinations of verb classes that we could use as "pointed verbs," we chose to evaluate these separately to find the set that was most accurate for classification and also to observe the effects of using stems. In this evaluation, we confirmed our intuition that the class "31.2 Admire Verbs" served our needs best.

Stemming

As this corpus of verbs only consisted of 45 verbs in one tense, it seemed clear that morphological stemming could be used to detect and use variants of these verbs in classification. A version of Porter's Stemmer (Porter 1980) was used to accomplish this.

Results and Coverage of Topical Sentiment Classification Approaches

As previously mentioned, I pruned our review corpus to the set of reviews that contained a literal mention of the topic name. Of this set, I used 5% for testing (847 reviews) and left the remaining 95% as part of the training data. The reviews in the testing set had a minimum length of 9 characters and a maximum length of 4682 characters, with an average length of 434.77 characters. Each classification method was attempted with and without domain knowledge, and on the full corpus of training data and on training data that only included topical sentences. The pointed verb classes were tested separately in a controlled experiment which determined that the "Admire Verbs" from the Levin Index were most accurate for use in sentiment classification.

In addition to accuracy, one other factor of the testing was important. For most methods, the goal was to limit the selection of words provided to the *RTS* system for analysis. In many cases, this limitation resulted in no candidate words to be scored for a testing document. As a result, the approach was unable to classify the document. For this reason, we kept a "coverage" count

for each method, that is, the number of testing documents that the method was able to classify at all. The end system must have full coverage in order to be useful in classifying blog content. The results of these tests can be seen in the following chart (Table 3-3).

Combination Approach

Upon analyzing the results, it was clear that the "Pointed Verbs Approach" provided the best classification accuracy. However, this approach was only able to classify about one eighth of the testing reviews. For this reason, I chose to implement an approach that iteratively combined the strategies of the most accurate methods. Since overall performance across approaches was best with the full training data set, not just the topical sentence training data set, I chose to use the full training data set in my final approach. This approach attempts to classify the target text using the "Pointed Verbs Approach with domain knowledge" first. If there is no result, the system then tries again using the "Adjectives in the Topical Sentence Approach with no domain knowledge." Finally, if the system still has no result, it uses the general *RTS* classification approach with domain knowledge.

This combination approach maximizes accuracy while having full coverage in classifying the testing data. Using domain information, and the full set of training data, this approach is 83.94% accurate, with full coverage of classification. Without domain data, the accuracy is still 83.47% with full coverage. Using domain information, and the topical sentence set of training data, this approach is 82.74% accurate, again with full coverage. Without domain data, the accuracy drops to 82.29%. This approach will suffice to accurately capture topical sentiment in text. The effects of this topical sentiment classification will be discussed in the following chapter.

Method	Accuracy with Domain Info	Coverage with Domain Info	Accuracy w/o Domain Info	Coverage w/o Domain Info
Full training data				
General full text classification	80.64	847	78.98	847
Adj in Sentence of Topic	82.47	673	85.29	673
Adj and Adv in Sentence of Topic	80.33	717	82.85	717
Adj and Verbs in Sentence of Topic	77.76	788	81.09	788
Adj, Verbs, and Adv in Sentence of Topic	77.68	794	81.49	794
Pointed Verbs	85.84	113	82.30	113
Pointed Verbs and Adj in Sentence of Topic	82.11	113	85.14	693
Combination Approach	**83.94**	**847**	83.47	847
2 word window around topic	78.66	836	77.39	836
3 word window around topic	79.08	846	78.25	846
4 word window around topic	79.31	846	79.55	846
Sentence of topic	78.63	847	79.57	847
Topical Sentence Training Data				
General full text classification	75.89	847	79.81	847
Adj in Sentence of Topic	83.36	672	84.67	672
Adj and Adv in Sentence of Topic	80.66	717	82.15	717
Adj and Verbs in Sentence of Topic	77.45	788	81.09	788
Adj, Verbs, and Adv in Sentence of Topic	76.85	794	81.11	794
Pointed Verbs	86.11	113	82.30	113
Pointed Verbs and Adj in Sentence of Topic	83.58	692	84.83	692
Combination Approach	82.74	847	82.29	847
2 word window around topic	75.06	836	77.51	836
3 word window around topic	75.95	846	78.37	846
4 word window around topic	75.98	846	79.20	846
Sentence of topic	76.57	847	79.81	847

Table 3-3 A table of accuracy and coverage for various approaches to topical sentiment classification.

CONCLUSION

In the context of enabling *Buzz* with the ability to distinguish between emotional and non-emotional stories, we've built a robust sentiment classification system that will scale to the needs of other systems. Our *RTS* system approaches human performance in the task of judging the emotional valence of a piece of text. Its major strength lies in the case based reasoning aspect which makes our system outperform similar systems by leveraging individual cases in order to preserve the discrepancies between emotional connotations of words across domains. A topic specific approach allows the system to focus on the sentiment surrounding certain topics in a piece of text. This approach was critical to the success of the topic based system described in the next chapter.

Chapter 4: Topical Stories, Opinions and Reactions

> *A great brand is a story that's never completely told. A brand is a metaphorical story that connects with something very deep – a fundamental appreciation of mythology. Stories create the emotional context people need to locate themselves in a larger experience.*
>
> - Scott Bedbury/Nike, Starbucks

Beyond digital theater installations, mining compelling stories from the blogosphere is powerful for many other reasons. Blogs provide people with a way to tell their stories in a public forum. These stories are often meaningful to businesses, as consumers are writing about their interactions with a product or service. However, given that millions of blog entries are written and posted daily, the challenge that remains is for businesses to sift through the blogosphere, digest the meaningful stories about their brands and products, and take action.

Buzz also exists as a way to expose compelling stories, reactions and opinions about brands and products. Unlike focus groups, stories and opinions found by *Buzz* are natural and not elicited. Watching a *Buzz* performance in this mode is useful not only as a virtual focus group, but also as a brainstorming tool for marketers because the virtual actors are presenting emotional stories about people's experiences with and feelings toward brands and products. Through

opinions expressed in a natural manner in the blogosphere, *Buzz* is able to connect companies to their consumers and to the general public.

To understand the need for such a system, telling stories about brands and products, consider a conversation that my advisor, Kris Hammond, and I had with the Director of IT at William Wrigley Jr. Co., Ian Robertson. Ian explained that in November of 2004, when Wrigley bought several candy brands from Kraft Foods, this purchase included the brand Altoids. Wrigley soon noticed that Altoids weren't selling very well in the United States. This was as expected considering that the candies are slightly untraditional as they often misshapen, messy and dusty, and sold in a tin that is awkward and has a paper liner inside. However, sales of Altoids in the Pacific Northwest were doing surprisingly well. So, Wrigley sent a marketing team up to Seattle where they stayed for 6 month, interviewing consumers and observing sales, until they came to the conclusion that Altoids were a grunge and highly sexualized product. Once this realization was made, of course, the marketing campaign was revamped nationwide, boosting sales greatly. While all this information was great, Ian wished that the team could have come to these conclusions in much less than 6 months. For this reason, we chose to focus *Buzz* on consumer stories about brands and products, in an attempt to provide marketers and business analysts with the necessary and immediate connections to their consumers.

TOPICAL STORIES

The story retrieval approach for *Buzz* (described in chapter 2) included using structural story cues such as "I had a dream last night" to seed the retrieval of stories. While the initial version of *Buzz* used topics (from Yahoo! Buzz and Wikipedia) to seed story retrieval, we found that it was difficult to formulate strategies and sources for automatically choosing topics that would be

of interest to a large audience. Topics such as popular actors, current events, etc. were often not engaging to the portion of the audience that was unfamiliar with this topic.

In order to find stories about brands and products, we returned to the notion of retrieval based on topics. While we previously had to focus on whether or not the audience would be interested in a topic, this is no longer an issue as we are certain that specific brands and products are exactly what interests the audience, a group of focused marketers or business analysts for a company. The focus no longer needs to be on finding interesting topics, but instead on finding interesting stories, reactions or opinions on provided topics of interest such as product and brand names.

Using all the standard filters developed for *Buzz*, the two most important features of topical stories is that the content is on point and emotional. For this system, relevance in topical stories is simply assessed via the count of occurrences of the topic name (or any deviations, provided as input) in the story. After a few passes with simple topical search using *Buzz*, we realized the importance of the emotional impact of the blog. Embracing this notion, we weigh the perceived emotional impact of a story more heavily than before, and more aggressively filter neutral stories. The result, as seen through three stories about Wrigley products in Table 4-1, was quite encouraging.

EMBODIED PRESENTATION OF STORIES

Wrigley Buzz was exhibited at the 2006 Wrigley Global Marketing Summit in Toronto from October 23rd to October 27th. The installation consisted of the central screen which displayed the focal topic words as opposed to the emotional words from the stories, as well as two avatars (above the central display) telling these stories about Wrigley products (product names were

provided by Wrigley). The actors took turns telling stories about a topic, such as "Juicy Fruit" (Figure 4-1), and then looked down to the central display, waiting for another topic to emerge (Figure 4-2). I was able to get feedback directly from Wrigley marketers at the Summit. They were both engaged in the presentation, as well as deeply interested in the idea of putting an ongoing permanent installation of *Wrigley Buzz* in a public space in the marketing department at their Chicago office.

It's been eighteen years since we started dating, and now we're married, we have a house, we have two kids. it's been good and bad, fun times and hard times, but I'm glad I picked you and you picked me. You are still my best friend, my closest confidant, the person who knows almost all of my secrets. I still like talking to you on the phone. The smell of Juicy Fruit gum will always remind me of you. I love you. Happy Anniversary!
my friends and family really made me happy, i'm so thankful to have them. it was alot of fun especially my gift from rae rae. rach decided to give me, handcuffs, candles, and cinnamon altoids, but i just dont know what i'm suppose to do with those three things.
but i can't help but be swamped by my few memories of him. him in his usual over-alls that he wore nearly everyday of his life. the wrigley's spearmint gum he always kept in the bib pocket and how he'd always pull my hair when i sat on the couch with him.

Table 4-1 Three stories about Wrigley products found and presented by the *Buzz* system.

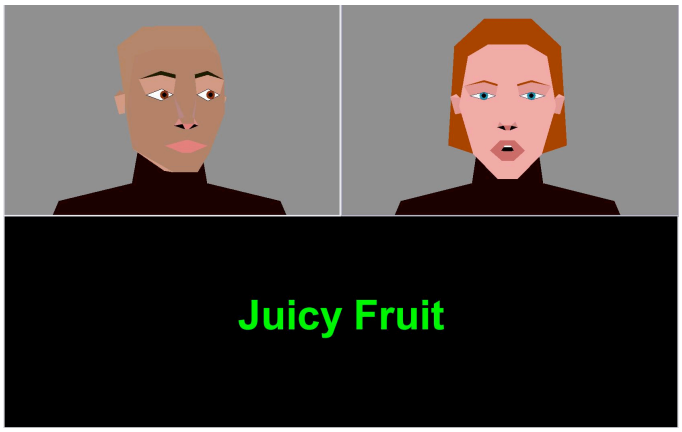


Figure 4-1 A single screen installation of Wrigley Buzz.

Figure 4-2 A single screen installation of Wrigley Buzz, in between topics.

While most of the feedback at the Summit was positive, I did receive some other suggestions or comments from the audience. A small number of viewers questioned the utility in embodying these bloggers with avatars and in having the avatars perform the stories. They cited various companies that provided graphical and numerical summaries and reports of the buzz around their products in the blogosphere (Cymfony 2007; NielsenBuzzMetrics 2007; Umbria 2007). While I understand the great utility of the tools provided by these companies, we take a different approach. We find the power of the story and of anecdotal evidence to be great as a brainstorming tool for marketing campaigns. These stories can be touching, funny, joyous, and

sad and we feel that their true impact can only be felt when they are told to the viewer by an embodied avatar.

Some theories of knowledge representation cite stories, also called cases, as the core representation schema in our minds. (Schank 1999) They explain that these stories are indexed in our memory by their details (locations, situations, etc.). In the following passage from *Dynamic Memory Revisited*, Schank explains the power of stories.

> *People tell stories because they know that others like to hear them, although the reason people like to hear stories is not transparent to them. People need a context to help them relate what they have heard to what they already know. We understand events in terms of events we have already understood. When a decision-making heuristic, or rule of thumb, is presented to us without a context, we cannot decide the validity of the rule we have heard, nor do we know where to store it in our memories. Thus, what we are presented with is both difficult to evaluate and difficult to remember, making it virtually useless. People fail to couch what they have to say in memorable stories will have their rules fall on deaf ears despite their best intentions and despite the best intentions of the listeners. A good teacher is not one who merely explains things correctly, but one who couches explanations in a memorable (i.e., an interesting) format.*
>
> - Roger Schank (Schank 1999)

This is not to say that charts and graphical reports provided by the blog analytics companies aren't useful. These charts can be very telling when temporally associated with events (product launchings, press releases, news articles, company hires and fires, etc.). However, we feel that to truly understand and internalize how consumers and the public feel about a product, stories are essential. Furthermore, presenting these stories in an interesting way is a critical part of

understanding. We feel that embodying the bloggers with avatars allows the viewer to more easily embed themselves in the narrative and internalize its impact.

> *Storytelling is the most powerful way to put ideas into the world today.*
>
> - Robert McAfee Brown

TOPICAL STORIES INVOLVING DRAMATIC SITUATIONS

Given that the system was successful in finding compelling or emotional stories about Wrigley products, we decided to take it a step further to detect not only topical stories, but topical stories that involve the dramatic situations that were compelling in the original *Buzz* installations. So, in addition to searching for blogs that included topic names, the system also looks for structural story cues, phrases that tend to indicate the beginning of a story. For example, one query could be: altoids "I had a dream last night…" In relevance evaluation, each chosen story must not only contain the topic name but also the structural story cue in order to be chosen.

This approach proved to be highly effective for topics that had a large presence in the blogosphere; a larger base of stories makes it more likely that the topic would co-occur in stories with these structural story cues. Table 4-2 is an example of three stories found using a combined approach that begins by looking for stories that include both the topic and a structural story cue. If no results are found, the system then relaxes the query to look for stories that only include the topic, but are highly emotional. The topics used were from a listing of Chicago landmarks, including topics such as Millennium Park, the Art Institute and Navy Pier.

OPINIONS

You might have noticed that the three Wrigley stories and Chicago landmark stories shown in Table 4-2 are all emotional stories, but don't exactly convey a direct opinion about the product, brand, or landmark mentioned. That is, the blogger didn't write "I'm sentimental about the smell of Juicy Fruit gum." Instead, she told an emotional story which conveyed how the smell of Juicy Fruit reminds her of her beloved husband. The blogger didn't say directly why they like or disliked Millennium Park, but instead told a story about a good experience that they had there. Such stories might not exist (or exist in a large volume) for all products or topics.

Dude, I went ice skating in Millennium Park, it's an outdoor ice rink, sort of like they have in New York. That was so wonderful. I am not so amazing when it comes to ice skating, which made me less excited about this adventure. But since it was a little bit of a dream of mine to skate downtown Chicago, with city itself the backdrop, I decided to go for it. I did much better than I thought I would. It was only my third time ice skating in my whole life, and I didn't fall down once.
Society tends to define human physical beauty by location and time period, but I think beauty is something each person sees in a different light. Besides, beauty is everywhere around you. I live in Chicago, and I find the architecture to be stunning here. The Art Institute is only a 10 min walk away, and it is filled with beauty. I can walk to the lake front and see the beauty of the water or go to Millennium Park and see how wondrous nature and manmade art can look combined.
I had a dream last night of walking to Navy Pier and playing in the square, marble fountain I use to love so much. Of course I couldn't. Classes may be over, but there are still tutors, study groups, notecards, and books demanding my every moment. All I could think about was that long walk to those great memories. It will happen soon enough. That's what I keep repeating to myself. It will happen soon enough.

Table 4-2 Three stories about Chicago landmarks, found and presented by the *Buzz* system.

Given my previous work that examined the familiarity of certain words and phrases, you might say that the system could easily judge what types of products are the ones that would be the focus of many stories and those which would not, by measuring their presence on the Web. However, many products or brands have a large representation on the Web, but what is written about them is generally uninteresting. This is because many products or brands (Ford, Nike, Panasonic, Kraft, etc.) play a role in our daily lives but we might not have a strong emotion or feeling associated with them. As a result, the product or brand names might be mentioned in blogs that take a more diary narrative form, where a product is mentioned in passing. For example, the following blog excerpts (found in a search on Google Blog Search for "Ford Taurus") contain the topical phrase "Ford Taurus" yet does not seem to convey any strong emotional state (Table 4-3).

"As we continued along the drive, we saw a sign that said, "Hamlin Beach Park". So we decided to stop at this park because it seems to be quite a popular place to stop by on the Seaway Trail. We got out of the Budget Rental Car (a Ford Taurus), walked across a large playground in front of the beach and then walked along the beach. For the most part, the beach felt extremely warm, however, the sand was somewhat hard. However, as you got closer to the the water, the sand felt softer and more loose." (http://blogswithaface.blogspot.com/2006_08_01_archive.html)
"In a press release acquired by T*he Phanatic*, the police indicated that the 24-year old Reid was operating a black Jeep Liberty that struck a red Ford Taurus, being driven by a unidentified 55-year old woman. Witnesses reported that Reid was traveling eastbound on Germantown Pike at a high rate of speed when he failed to stop at a red light on Arch Road. Reid's vehicle subsequently struck the rear driver's side of the female's vehicle." (http://daily.phanaticmag.com/2007/01/garrett-reid-in-possession-of-drug.html)

Table 4-3 Two blog excerpts containing the phrase "Ford Taurus," yet not giving any opinion or emotion towards "Ford Taurus."

For most products, if emotional stories about a product do exist, they are likely embedded in a larger corpus including narratives of the type described and shown above. For this reason, we sought to mine more than only stories from this corpus, extending the *Buzz* story finding

capabilities to function as a way to find opinions on a topic. We use the structural story cue approach in *Buzz* to seed the system to look for things that look more like opinions than stories. For example, when a blogger uses an opinion phrase such as "I think," "I feel," "I believe," "I love," "I hate," or "my favorite" in the same sentence as the topic, then it's likely the blogger is stating an opinion about the topic.

For opinions, sentiment plays an even larger role. We want to find the opinions that are emotionally amplified. More so, we want to find opinions where the blogger is emotional about the specific topic of focus. For this reason, the topical sentiment strategies, described in the last chapter, become critical for *Buzz* in finding opinions about topics. That is to say, if a story is generally highly emotional, but not necessarily emotional about the topic at hand, then the story might not be interesting to the viewer.

Three resulting opinions about Ford Mustangs, found by *Buzz* using the techniques described, are shown below. It became clear to us very quickly that searching explicitly for opinions was a much more powerful way to find stories or opinions that would be interesting or useful to marketers across a broader range of brands and products.

I love Ford Mustangs. I think this comes from my encounter with that really hot guy on the highway a few months back. Ever since I've become really interested in variations of this car. My favorite kinds are the ragtop convertibles. It's the original American Babe Magnet. If you're a hot guy that drives a Mustang, it should be guaranteed that you'll have babes all over you.
Just got back from watching this Al Gore movie about Global Warming. I've never been a big fan of his as a political leader. He does have a passion for this subject and I really respect it. It just makes good sense for us to do what we can to reduce C O 2 emissions now, before it is too late. Around our house, we try. We recycle glass, plastic and paper. We bought an energy efficent refrigerator, washer, dryer. I'm struggling about the car thing. We have a 2004 Volvo Cross Country we bought in February and my 2001 Ford Mustang Convertible. We need one larger car. I love my convertible, but it gets pretty lousy gas mileage for such a small car. Can someone make a hybrid convertible?

I was on my way home, and my friend told me there was a Ford Mustang driving behind me. That's like my favorite car in the world. So I pull over and let him pass me, only to continue and follow him through this maze of streets, somewhere in the suburbs. He finally parks the car, and as I pass him very slowly, he looks up at us with this scared look in his eyes. So, we waited on the corner till the guy was inside his house, turn back and check out the car. I want it!! I want it!!

Table 4-4 Three stories about Ford products, found and presented by the *Buzz* system.

CONCLUSION

In the end, *Buzz* is able to distill compelling consumer stories and present them to marketers and business analysts in an engaging manner. Through the filtering, modification, and retrieval model, the system extracts poignant stories of consumer experiences with and opinions about focused brands and products. Supplementing traditional marketing research, *Buzz* is often able to find stories and perspectives that were previously only accessible through consumer interviews and focus groups. The stories are not only useful for digesting and understanding consumer feedback, but as a tool to brainstorm for marketing campaigns. Presenting these stories through avatars with computer generated voices; the result is an engaging user experience that provides marketers with a connection to their consumers through stories.

Chapter 5: Compelling News Stories and Perspectives

> *It's all storytelling, you know. That's what journalism is all about.*
>
> - Tom Brokaw

Living in a world so connected via the internet gives us the ability to access many different points of view on a current event, no longer relying solely on the often one-sided story presented by the nightly news anchors. Given access to this vast amount of opinions, digesting these opinions is not as easy as it would appear.

News at Seven (Nichols, Hammond et al. 2006; Nichols, Owsley et al. 2007) is a completely automatically generated news show presented through a realistic modern gaming engine. It is complete with digital anchors in a virtual set, interviews with the public, and background footage of the stories it presents. The end goal of this system is to move away from broadcast news in two dimensions. First, provide a completely customized and personalized news show produced on demand. Secondly, the system presents a variety of automatically mined opinions on news stories from the blogosphere. They represent the points of view of bloggers whose opinions on these current events are emotionally amplified. For this system, I modify the techniques used by *Buzz* to now find compelling and emotional opinions as opposed to stories. Similar to *Buzz*,

News at Seven is able to connect people via opinions and points of view, as expressed in the blogosphere.

The script generation system for *News at Seven* has three major components. The first involves the actual news story itself. It's important to note that *News at Seven* doesn't actually "write" the news, but gathers and alters existing online news stories while changing the content to be more appropriate for a spoken news show. Second, the system gathers media, specifically videos and still images, to supplement the anchor and reporters. Finally, the system finds alternate points of view for the news story. These points of view are pulled from blogs that are relevant to the story and have highly affective content. While three components are all quite important, my focus is on the latter.

FINDING OPINIONS RELATED TO NEWS STORIES

Given the connected nature of *News at Seven*, providing alternative points of view is essential. Millions of people are posting their opinions daily in the form of weblogs (blogs), yet standard broadcast news shows allow us to consume only one side of the story. To find alternative points of view, *News at Seven* mines the blogosphere for relevant and emotional points of view related to the story at hand.

In chapters 2 and 4, I presented *Buzz*, which mines the blogosphere for compelling and emotional stories about a brand or product, or just generally compelling stories (dreams, fights, nightmares, confessions, apologies, etc.). *Buzz* can be seeded with a topic (a word or phrase) and/or phrases that indicate the beginning of a story such as "I had a dream last night…" or "I have to confess…" *Buzz* retrieves blogs using Google Blog Search given the seed as a query. It iterates through the results and identifies candidate stories (excerpts from blogs) which are on-

topic to the given seed. The candidates are then passed through a set of filters to remove those which do not appear to be in story form. For example, those with punctuation that is too dense are removed as they appear to be a list or survey. Those which contain profanity or language that refers to the act of writing or blogging may be removed. The remaining candidates are then scored based on their relevance to the seed, as well as by their emotional content, using an emotional classification system that uses a combination of case based reasoning and machine learning and is trained on movie/product reviews (Sood, Owsley et al. 2007). The result is a complete story, in paragraph form, that is emotional and relevant to the given seed.

In order to target opinions about news stories, we utilize the *Buzz* structural story cue approach to seek out stories that begin with a phrase that indicates that the writer is about to give an opinion. This includes phrases such as "I think," "I believe," "I feel," "I can understand," "I appreciate," etc. Using these phrases as seeds, we are likely to get stories that have the feel of an opinion or point of view.

To find blogs that are not only opinions, but are also relevant to the news story at hand, we assess and re-rank the results based on their relevance to the news story. Given a news story, we form a query based on its content. This query is then sent off to Google Blog Search (Google 1996), which returns a set of candidate blogs. Next, these blogs go through the series of filters built for *Buzz*, which remove content that is inappropriate or uninteresting as a story. A few additional filters were created for this system, to meet the needs of finding current event opinions as opposed to stories. With the set of stories that remain, relevance scores are calculated, along with emotional impact scores. Since we are concerned with the opinions being both on point and emotional, re-ranking of the opinions from blogs then involves using a weighted average of the relevance and emotional scores.

FORMING QUERIES

In order to find opinions related to a news story, we must first distill the story down to its salient terms, to be used as a query to a search engine. To form a query from a news story, the system utilizes both a statistical model and an emphasis model of the document in order to determine which words are most important. Since the source document is a news story with a title, the system perceives the title as more important than the body of the news story, as the title is intended as the shortest possible summary of the story. Throughout the story, any terms that are capitalized or bolded are also seen as more important than lowercase and non-bolded text, as the author is likely to have bolded the words with the intention of emphasizing them. In combination with this model of emphasis, the system uses a statistical model of the document in order to determine the most salient terms in order to form a query.

For the statistical representation, the body and title are combined to form a document, which is then analyzed to form a *tfidf* representation. From this representation, the top 20 words (sorted in decreasing order by *tfidf*) are examined. If any of these words appear starting with a capital letter in the title, or share a common stem with a word that starts with a capital letter in the title, then the word is automatically added to the query term vector. The remaining words in the query are chosen based on the highest *tfidf* score, giving preference to those terms that started with a capital letter in the news story, where all terms used in the query must be above a minimum *tfidf,* above a minimum *tf* (term frequency), and comprised of more than two letters. This model emphasizes the importance of words in a news article title, as well as words that are capitalized in the story, likely proper nouns.

Using Google Blog Search, we found that a query is six terms was ideal; balancing between getting too many results and getting no results. Once a topic query is formed, multiple queries are formed by combining the topical query with various opinion starters such as "I think," "I feel," or "I believe." Each query is sent to Google Blog Search. If there are an insufficient number of results, the final term of the query will be dropped and the query will be sent again. This process continues until a sufficient number of results are found, or until the query only contains one term.

ASSESSING RELEVANCE OF CANDIDATE OPINIONS

In many cases, search engines do all that is possible to evaluate the relevance of the results to a given search query. However, in this case, our system has more information than the query itself; it also has the original news story from which the query was formed. To find the most relevant opinion, we utilize more than just the query formed, but also the entire text of the news story. Since we have the original new story as a context, and the text of each opinion candidate (blog excerpt), we are able to re-rank these results returned by the *Buzz* engine, by using a document similarity comparison common in the field of Information Retrieval (Salton 1983; Salton and Buckley 1988).

To begin this re-ranking, we create a *tfidf* representation of the original news story. This statistical representation exposes the salient terms in the document, taking into account the frequencies at which words occurred in the document, compared to how common the words are in a large corpus of documents. The representation consists of a list of words from most important to least important (largest to smallest *tfidf* score) where each word has an associated

frequency (frequency of occurrence in the document), document frequency (frequency of occurrence in the large corpus), and *tfidf* score.

The *tfidf* score for a word *i* in a document *j* is calculated by the following equation:

$$TFIDF_{ij} = \frac{frequency_{ij}}{\max frequency_{j}} * Log\left(\frac{N}{n_i}\right)$$

Equation 5-1

where $frequency_{ij}$ denotes the number of times the word *i* appears in the document *j*; *N* is the number of documents in the larger corpus, and n_i is the number of documents the word *i* appears in within the corpus.

After we generate the *tfidf* representation for the news story, we then create the same representation for each of the opinion candidates found by the *Buzz* engine. Given these representations, the system calculates a relevance score for each document. To calculate this relevance score for an opinion candidate, the system iterates through the words in the *tfidf* representation of that opinion. For each word, if it exists in the *tfidf* representation of the news story, we make an addition to that opinion's relevance score. This addition involves creating a weight factor of the word based on its rank in the *tfidf* representation of the news story (with the highest ranked word having the largest weight factor). The addition to the relevance score is then equal to the summation over *k* from 1 to *frequency* (the number of times the word occurs in the blog excerpt) of the weight factor of the word multiplied by 1 divided by the sum of k and 1. The opinions are then re-ranked by their relevance scores, calculated as follows:

$$\sum_{i=0}^{n} \sum_{k=0}^{frequency_{ij}} Weight_i * \left(\frac{1}{k+1} \right)$$

Equation 5-2

Using this approach, the system is able to evaluate the relevance of the opinions found by the *Buzz* engine, allowing us to present the most relevant emotional opinion.

ADDITIONAL FILTERS

In addition to the standard *Buzz* filter set, some additional filters were needed in order to deal with the types of opinions or candidate stories that were found using this method. Since we are looking for opinions relevant to an online news story, it's clear that the most relevant opinion would be one completely comprised of direct quotes from that news story. For example, the following is a candidate opinion found by the *Buzz* engine in relation to a news story about the West Bank:

> *"We have always asked for international forces to come to the West Bank and Gaza," Abbas confidant Saeb Erekat told Israel's Army Radio. But, he added, "Honestly, on the personal level, I believe that if we don't help ourselves as Palestinians, nobody can."*
>
> - AggressiveResponse (AggressiveResponse 2007)

While we know that this is relevant to the news story, it's not an opinion, and more importantly, it's not interesting in the context of a presentation. To handle this, I created an additional filter to remove candidate stories that are comprised of more than 30% direct quotes. This allows for

some quoting of the news story, but requires that the majority of the opinion or story is non quoted material. In addition, if the phrasal opinion indicator ("I feel," "I believe," "I think") falls within a direct quote, then the candidate opinion is also filtered out as the blogger is not giving an opinion of their own, but conveying someone else's.

Figure 5-1 A still shot from a *News at Seven* show, depicting an opinion segment.

OPINIONS ON NEWS STORIES

This system's performance in finding emotional opinions relevant to news stories has been quite promising. Below you'll find a set of news story excerpts paired with related opinions found by this system (Table 5-1). In a *News at Seven* show, these excerpts are spoken by a "man on the street" to show that it is the opinion of an average viewer (Figure 5-1).

News Story	Related Opinion
Title: Weak `Idol' wannabe gets online support **URL:** http://abcnews.go.com/Technology/wireStory?id=2988365&CMP=OTC-RSSFeeds0312 **Story:** Flowing hair and a precious smile have their rewards. Especially if you're Sanjaya Malakar, who is considered one of the weakest performers on "American Idol" but has a fan base that has helped him survive multiple rounds of viewer elimination. In the online community and in Malakar's home state of Washington, the croaking crooner seems to have a loyal following of friends, family and fanatics who would like nothing better than to see him achieve the ultimate "Idol" success and be the last singer standing in May.	**URL:** http://bumpshack.com/2007/03/27/american-idolone-more-must-go/ More of the usual this week on American Idol. The same ones shined and yet another queer hairstyle from Sanjaya Malakar. The Bumpwife immediately said that boy needs a therapist. You know I think she is right. She actually use to be in his camp, but her jaw was on the floor with Sanjaya came out.
Title: Iran softens stance on British sailors **URL:** http://news.yahoo.com/s/ap/20070327/ap_on_re_mi_ea/british_seized_iran_56 **Story:** Iran said Monday it was questioning 15 British sailors and marines to determine if their alleged entry into Iranian waters was "intentional or unintentional" before deciding what to do with them — the first sign it could be seeking a way out of the standoff.	**URL:** http://journals.aol.co.uk/pharmolo/NorthernTrip/entries/2007/03/25/iran-and-iraq/4047 I think it is important to keep an eye on the utterances of Mr Ahmadinajad, even if I don't like what he says. I am acutely aware of the increasing tensions in his part of the world, not helped by the plight of the 15 British naval personnel who strayed into disputed waters to investigate an Iranian boat. The sailors were apprehended in a part of the Persian Gulf that is claimed both by Iran and Iraq. I somehow don't think the British navy was aware of that. It has handed the Iranian hard-liners a perfect excuse for inflaming the situation even further.
Title: Prosecutors: Revoke Hilton's probation **URL:** http://news.yahoo.com/s/ap/20070330/ap_en_tv/paris_hilton_17 **Story:** City prosecutors said Thursday they will ask a judge to revoke Paris Hilton's probation in a reckless driving case, a move that could lead to a jail term. The decision followed an investigation into whether the hotel heiress and reality star violated terms of her probation by driving last month with a suspended license. "We're confident we have sufficient evidence to prove that her license was suspended and that she had knowledge of that suspension," said Nick Velasquez, a spokesman for the city attorney's office. He declined to elaborate on the evidence, citing an ongoing investigation.	**URL:** http://zanesvilletimesrecorder.com/news/blogs/stalepopcorn/2007/03/prison-blues-for-paris-top-april-fools.htm Paris Hilton may be headed for the slammer. LA prosecutors are asking that her probation be revoked, and jail time is a possibility. I think the streets of America would be safer if Paris wasn't on them.
Title: Court says `Da Vinci Code' not a copy	**URL:**

URL: http://news.yahoo.com/s/ap/20070328/ap_on_en_ot/britain_da_vinci_code_10 **Story:** Britain's Court of Appeal rejected a lawsuit Wednesday from two authors who claimed novelist Dan Brown stole their ideas for his blockbuster novel "The Da Vinci Code." Michael Baigent and Richard Leigh had sued Brown's publisher, Random House Inc., claiming he copied from their 1982 nonfiction book "The Holy Blood and the Holy Grail." Both books deal with the theory that Jesus married Mary Magdalene and had a child, and that the bloodline continues. One of the judges said copyright protects an author's labor in researching and writing a book, but does not extend to facts, theories, and themes.	http://lifeouteast.blogspot.com/2006/12/holy-blood-holy-grail-da-vinci-code.html About fifteen years ago I read Holy Blood Holy Grail by Michael Baigent, Richard Leigh, and Henry Lincoln. It's an historical detective piece based on research and fact and a lot of gossip and myth. However, they put forward an interesting viewpoint that is, in parts, highly believable. Dan Brown's The Da Vinci Code has got to be at worst a blatant ripp-off and at best heavily inspired by Holy Blood Holy Grail. I think he won the court case against him but how anyone could fail to see the connection and striking similarities is mind boggling.

Table 5-1 Four sample *News at Seven* stories and the related opinion found by *Buzz* and presented in the *News at Seven* opinion segment.

CELEBRITY NEWS

Given the vastly different types of stories that live in the realm of "news," it is clear that different modes of presenting the news must exist. For example, the fast paced and upbeat music that characterizes an "Access Hollywood" type show would not be an appropriate medium to tell a story about the war in Iraq. While these distinctions are especially important for presentation, they also have an impact on the types of alternative points of view that the system gathers in creating a script. For a standard world or political news story, *News at Seven* uses an opinion from a blog found using the method described above. That is, the blog is directly relevant to the news story and contains a phrase such as "I think" to indicate that it is an opinion. The opinion is presented by a "man on the street," as depicted in Figure 5-1.

However, for entertainment news, the dynamic can be quite different. To increase the pace and change the presentation dynamic for an entertainment story, *News at Seven* uses an

interruption model. Instead of presenting an opinion at the end of the story, a second anchor gives short opinions about every celebrity mentioned, as they are mentioned by the anchor. For example, the following is a snippet from a *News at Seven* Celebrity Edition script, where **Anchor 1** is presenting a celebrity gossip story, and **Anchor 2** is presenting emotional opinions about these celebrities, found in the blogosphere by *Buzz*.

Anchor 1: Hollywood's star couple Brad Pitt

Anchor 2: Brad Pitt is desperate for a biological son.

Anchor 1: and Angelina Jolie

Anchor 2: Angelina Jolie is mesmerizing.

Anchor 1: are all set to add an extra glitz to the Cannes Film Festival this year.

These short opinions were all found using the *Buzz* system. The system was configured to look for one sentence opinions, under a certain length, that contain the celebrity's name and are highly emotional. The system is quite effective in finding opinions that meet these criteria. For stories in the realm of entertainment, this model works well as the opinions that bloggers general write in this area are shorter and less deep than opinions in say, the political arena. More examples of short celebrity opinions found by *Buzz* in this mode are shown in Table 5-2.

In general, different types of news stories have different types of opinions that are most relevant to them. For general news, it is long winded opinions about the topics discussed in the piece; for celebrity gossip, it's opinions specifically about that the celebrities mentioned. As we continue to develop other types of segments, we expect to adapt our blog-retrieval strategies to meet both the presentational needs, as well as adjust to the types of opinions that are expressed in that domain.

Celebrity	Related Opinion
Angelina Jolie	Angelina Jolie is mesmerizing.
Brad Pitt	Brad Pitt is desperate for a biological son.
Vince Vaughn	Vince Vaughn is not attractive at all.
Jennifer Aniston	Jennifer Aniston is my favorite actress.
Madonna	Madonna is actually ok, other than being a total joke.
Guy Ritchie	Guy Ritchie is brilliant at what he does.

Table 5-2 Six celebrities and the related opinions about them from the blogosphere, found and presented in a *News at Seven* Celebrity Edition show.

CONCLUSION

The concept of *News at Seven* alone is quite exciting and promising; as an automated news show production system, it could revolutionize how we hear about and see the events of the day. Imagine the possibilities of being able to automatically mine opinions that are not only relevant to a news story, but emotional, evocative, and controversial, and yet still within the range of what is appropriate for the news story type. Through a system like *News at Seven*, we can present not only news, but points of view, connecting people to people through emotional responses to shared events in the world around us.

Chapter 6: Network Arts and the Association Engine

> *In Network Arts, the installation is the mediator between people and culture. It is the job of the agent and the artist together to deliver an emotional message to the audience.*
>
> - David A. Shamma (Shamma 2005)

Network Arts (Shamma, Owsley et al. 2004; Shamma 2005) is an area focused on the creation of informative and compelling performance installations that find, use and expose the massively interconnected content that exists on the Web. Systems in this area both inform viewers of the current state of the Web, and enlighten them with associations and points of view. These systems are important because they bridge the gap between the internet and digital art, an area that traditionally uses the machine as an instrument which is utilized by a human artist. Using the Internet as a resource, systems in this area are scalable and autonomously generate creative and entertaining experiences that provide a reflection of the world that we live in.

The first system in Network Arts is an installation called the *Imagination Environment* (Shamma and Hammond 2004). This system amplifies media viewing experiences such as movies, music videos, or news, by presenting the viewer with imagery to supplement the media, displayed on eight surrounding screens. The images are meant to represent possible associations in one's imagination to the media, as a series of associations from words (spoken in the media stream), to images found on the web. This work has been quite successful and exhibited in many

public installations including Chicago's Second City Theater, Wired Magazine's Next Fest, and was reviewed in the New York Times. It exposes the associations between words and images on the web, but such associations are grounded in the context of a movie, music video, or television news.

The *Association Engine* followed the *Imagination Environment*, as an effort to create a troupe of digital improvisers (Owsley, Shamma et al. 2004) that could autonomously generate a creative and entertaining experience. A team of five digital actors, with animated faces and voice generation, could autonomously perform a series of improvisational warm-up games, followed by a performance. This system is important as it was a stepping stone to my work on *Buzz*, approaching the problem of how to build a system that can generate a compelling and entertaining experience.

> *Good improvisation isn't thinking about those things. It's finding your individual deal with another's individual deal and realizing a common context and surprising from within it. Plain and simple.*
>
> - Mick Napier (Napier 2004)

THE ASSOCIATION ENGINE

While there are many published guidelines of improvisational theater, many of the great improvisers say that they don't follow these rules. Improvisation is about connecting with and reacting to both the other actors and your audience. It is largely about the associations that each actor has to words and phrases, which are based on their own life experiences.

It's hard to imagine how creating a digital improviser would be possible. How can a system embody the experiences and associations from one's life, and access them? How could the

system's experiences grow in order to provide novel associations? How could it scale to represent different personalities and characters?

The Association Engine began as an attempt to create a digital improviser and in the end a troupe of digital improvisers that could generate a creative and entertaining performance. *The Association Engine* includes both a backend system which scripts the actors as well as a front end embodiment of the actors. Both sides of the system are important and are described in this chapter.

THE PATTERN GAME

In Improvisational Comedy, troupes generally gather before performances to warm up, and get on the same contextual page. There are a variety of ways that troupes do this. One common way is a game called the "Pattern Game," also known as "Free Association," "Free Association Circle," or "Patterns." There are many variations to this game, but there are some very basic rules that are common across all variations. The actors stand in a circle. One actor begins the game by saying a word. The actor to the right of this actor says the first word that comes to their mind, and this continues around the circle. The actors try to make contributions on a strict rhythm to ensure that the associations are not thoughtful or preconceived. Some variations of the game encourage the actors to only associate from the previous word, while others require that the associations are in reference to all words contributed so far. In some cases, the actors attempt to bring their associations full circle and return to the original word. The goal of all variations of this game is to get actors warmed up on the same contextual page and in tune with each other before a performance.

Our first step in creating a digital improviser was the modest goal of creating a system that could participate in the "Pattern Game." If we are able to create a digital improviser that can participate in the "Pattern Game," then we can build a team of improvisers which can generate a shared context, and eventually do a full performance. In order to do so, we began by enabling the system with access to some set of possible associations to words, as well as a method for choosing amongst them. There are many available corpora of word associations published through thesauri. Of these, Lexical FreeNet and WordNet are the largest accessible online connected thesauri.

Connected Thesauri

WordNet (WordNet 1997), developed by a group of researchers at Princeton University, holds 147,249 unique noun, verb, adjective and adverbs. For all words, the system provides synonyms divided by the senses of the word they are similar to, antonyms when appropriate, and a word familiarity score. For each noun, the system also gives hypernyms, hyponyms, holonyms, and meronyms. For each verb, troponyms are also provided.

While WordNet is a rather large thesaurus, Lexical FreeNet (Beeferman 1998) is nearly twice its size. In addition to having many more words, 256,112 unique tokens, the system not only shows relations between words, but also concepts and people. The relations that it provides are more wide ranging including: "Synonymous," "Triggers," "Generalizes," "Specializes," "Comprises," "Part of," "Antonym of," "Rhymes," "Sounds like," "Anagram of," "Occupation of," "Nationality of," "Birth year of," "Death year of," "Bio triggers" and "Also known as."

Given that Lexical FreeNet provided more relations per word and a larger word set, we chose to use this connected thesaurus as a word repository for the digital improviser. Using this

corpus as a source of association, the agency must have a method for choosing their contribution from the set of possible related words. In building this mechanism, various approaches were attempted and the successful strategies were combined in the final system.

Word Familiarity

First, many of the words and associations in Lexical FreeNet are very obscure. For example, in Lexical FreeNet, there are 508 words related to the word "cell." Included in this set are "cytoplasm," "vacuole," "gametocyte," photovoltaic cell" and "bozell." In human improvisation troupes, actors would not contribute a word like "gametocyte" to the Pattern Game for a few reasons. They are warming up with the intent of generating a context from which to do a performance. Because this is aimed towards a future performance, they will not use words that would be unfamiliar to their audience as this would result in the audience becoming disengaged. Just as we use vocabulary that is familiar to someone we are engaged in a conversation with, the content of a performance must be familiar and understandable to the audience. Secondly, they would not make associations that the other actors might not understand as that is counter productive to the goal of getting them on the same page. An actor can't be expected to free associate from a word they are not familiar with. Similarly, overly common words are not advantageous as they are generally uninteresting, and don't provide a rich context for a show.

For these reasons, we sought to enable the digital improvising agency with the ability to avoid words that are overly obscure or too common from the related word set provided by Lexical FreeNet. While WordNet provides a familiarity score for each word, it did not appear to us that these scores gave an accurate reflection of how commonly the word is used. To generate an accurate measure of familiarity, we looked to the web as an accessible corpus of language use.

For any single word query, doing a web search through a modern search engine provides us with a number of results. We chose to use this number of results as a measure of the familiarity of the word. Not only does this give us a measure of familiarity, but it allows for change over time because when new concepts and words come into common use through current events, scientific advances, etc., their presence on the web will increase, thus increasing their perceived familiarity to our system. For example, prior to November of 2002, the acronym "SARS" was very uncommon. However, after the outbreak of "SARS" in mainland China, the word was very commonly known in the public as it was reported in countless news stories and blogs. As its web presence increased, so did its familiarity to the general public.

To create thresholds for minimum and maximum familiarity of an acceptable association, we needed a corpus of words to examine a distribution of word frequencies. To create this corpus, we took a one day snapshot of 14 Yahoo! News RSS Feeds. From these feeds, we gathered all the news stories. We processed and tokenized these stories, generating a list of all 4,201 unique tokens that occurred. After this list was completed, we sent each unigram to Google as a single word query and captured the quantity of results that Google returned for each query. The query term with the most results was "the", with 5,240,000,000 results, while "pseudo-machismo" had the 11 results, the least when ignoring all tokens with zero results. Two graphs of these frequencies are shown below; the bottom graph shows the frequencies on a logarithmic scale for readability.

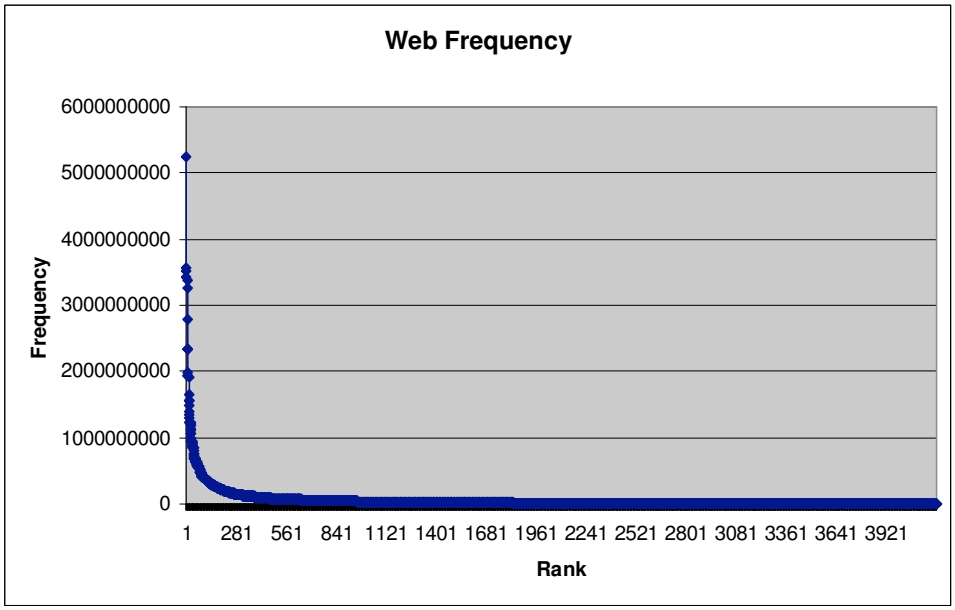


Figure 6-1 A graph of words and their frequency of occurrence on the web.

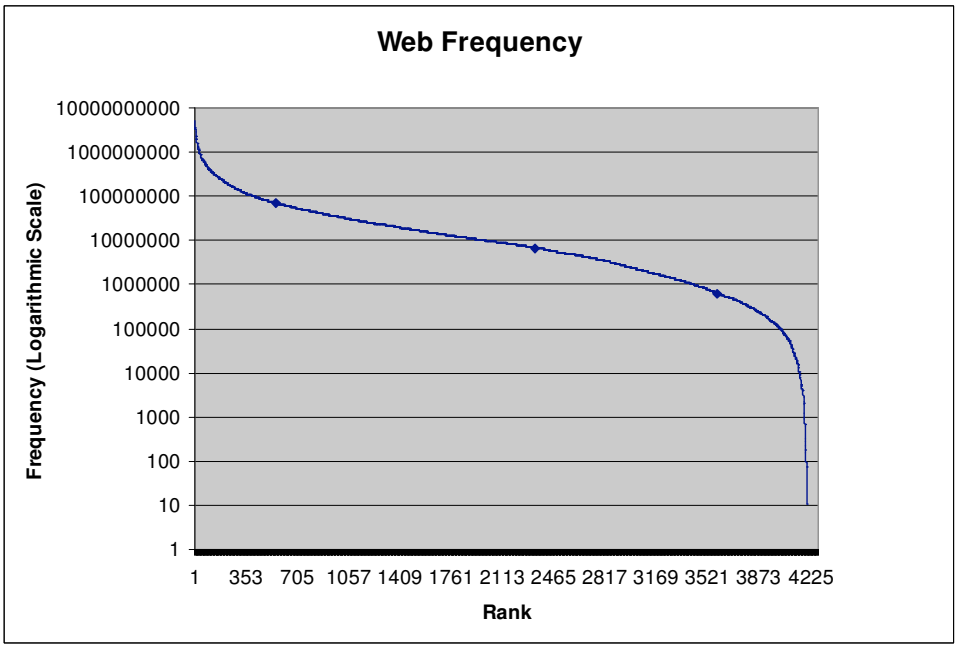


Figure 6-2 A logarithmic graph of words and their frequency of occurrence on the web.

From this distribution, we generated the minimum and maximum thresholds as one standard deviation from the average frequency. The maximum threshold was 69,400,000 and the

minimum threshold was 654,000. The digital improvisers then consider the familiarity of their word choices by only choosing words which fall between 654,000 and 69,400,000 pages on Google.

To validate this tool as an accurate judge of word familiarity, we conducted a study with 202 participants. The set of 4,201 unique words above were divided into six 15% percentile frequency range groups. For this study, words were chosen randomly from six percentile groups, ignoring terms with a frequency greater than the 83rd percentile (beyond one standard deviation) as these words were common "stop words" for which we had standard methods of ignoring. To avoid the affects of priming, the words were presented to the participants in random order. The participants judged each word on a scale from 1 to 5, unfamiliar to familiar.

Dependence of the Google frequency on the participant's data was found ($p = 0.9130$). As a comparison, WordNet's familiarity measure showed no dependence on the participant's data ($p = 0.0014$). From this study, we concluded that Google frequency was indeed a good measure of word familiarity.

Context

As mentioned previously, there are several different varieties of the Pattern Game. We chose to implement a version where the actors associate not only from the previous word, but from the context of all previous words being contributed. This keeps the actors on point, and tied into a space of words. When one word space is exhausted, they can jump out of it with an association into a different space or set of words. The ending result is that the team has one or multiple clear topic areas within which they will do their performance.

To emulate this behavior within our digital improvisers, we use a sliding window of context. Contributions are chosen not merely from the set of words related to the previous word contributed, but from the intersection of the sets of words related to the last *n* words contributed, where *n* is decreased when the intersection set of related words is sparse or empty. This method resulted in selection of words that stays within a context for some time and then jumps to a new word space when the context is exhausted, much like how human improvisers perform in this game.

Relation Types

To maintain novelty and flow in the pattern game, human improvisers will not make redundant associations. For example, six rhyming words will not be contributed in a row. Conversely, some improvisers might lean towards particular relation types. For example, an actor might contribute antonyms whenever possible.

To take these two characteristics into account, the digital improvisers use memory of previous relations and tendencies to guide their decisions. Remembering the previous *n* associations made, they can avoid those relation types where possible. They can also be seeded with tendencies towards particular types of relations, “kind of,” “synonym,” etc., using these relationship types whenever possible.

The Pattern Game Generated by Digital Improvisers

The final backend system is one that uses all the methods described above in order to choose a related word to contribute to the Pattern Game. The system first takes a seed from the audience through keyboard input. This ensures that the system is live and not running cached

demonstrations. To make a contribution, the digital improviser first finds the intersection set of the sets of related words to the previous *n* words. Then, from that set, it eliminates those words which are too familiar or too obscure. It then takes into account its own tendencies towards relation types and the recent history of relation types used in order to choose a word to contribute to the game.

Here is an example string of associations made the digital improvisers given the input word "music." "Music, fine art, art, creation, creative, inspiration, brainchild, product, production, magazine, newspaper, issue, exit, outlet, out."

Here is a second example, starting with the input word "student." "Student, neophyte, initiate, tyro, freshman, starter, recruit, fledgling, entrant, beginner, novice, learner, apprentice, expert, commentator."

ONE WORD STORIES

Improvisational games and performances can take many different forms. A common game is the "One Word Story," also known as "Word at a Time Story." To do this game, the troupe again stands in a circle. One actor starts by saying a word to begin the story. Moving around the circle, each actor contributes one word at a time until the story is complete. At the end of sentences, one actor may say "period." Like any other performance, this game is usually done after a warm-up so that the troupe is on the same contextual page from which the story can be told. While simplistic in interaction, this game is surprisingly hard for new actors.

Our next step in building a digital improviser was creating a team that could participate in and create a compelling performance of the "One Word Story" game. To do so, we used a

template based approach, choosing and filling templates based on the resulting context of the pattern game.

Generating Templates

Taking a template based approach to story generation; we first generated a library of story templates which indicate how different types of stories are told. For this system, we chose to generate stories similar to the types of stories in Aesop's fables (Pinkney 2000) as they are short and simple, yet still have a moral or lesson. We generated a set of twenty-five story templates, somewhat similar to the children's word game "MadLibs" (Price and Stern 1974). The goal was to be able to generate stories which were both original or novel and interesting. This was done by making the templates simple, with parameterized actors, locations, objects, and emotions.

Below are four of the twenty-five parameterized templates used by the system. The types of each blank or parameter for the story are defined above each story. For example, in Story Template #1, the system must fill in the blank labeled "<0>" with a "female-name." This name will be used again throughout the story whenever the "<0>" is referenced. While games such as "MadLibs" reference the parameters by parts of speech and the like, we found that more specific parameter types could result in a more coherent story.

Story Template #1

```
#       0       female-name
#       1       employee
#       2       employee
#       3       building
#       4       emotion
&
There once was a woman named <0>.  <0> wanted very much to be a <1>  ,
but no one thought she could do it .  To become a <1> , <0> went to
the <3> , where all of the <1> people gather .  Unfortunately when <0>
```

got to the <3> , she found out that all of the <1> people had become
<2> people. <0> felt <4>.

Story Template #2

```
#       0       male-name
#       1       material
#       2       school-subject
#       3       tool
#       4       material
&
```

<0> was taking a class in <2> . For his <2> class , <0> had to build a
project . <0> had planned to use a <3> to build his project out of <1> .
It turned out that his <3> did not work on <1> , so he had to use <4> instead.

Story Template #3

```
#       0       employee
#       1       female-name
#       2       female-name
#       3       show
#       4       hall
#       5       hall
&
```

A <0> named <1> called her friend <2> . <1> wanted to go to the <3> at the
<4> . <1> and <2> met up at the <4> . To their surprise there was no <3>
at the <4> . Instead, the women decided to go to the <5> .

Story Template #4

```
#       0       male-name
#       1       walk
#       2       animal
#       3       animal
#       4       weather-condition
&
```

<0> began to <1> through the park one day and came across a wounded <2>.
A <3> was near and helping the wounded <2>. The weather was <4> , so
<0> continued to <1> and left the wounded <2> in the park.

Table 6-1 Four sample story templates used by *The Association Engine.*

One important feature of these templates is the notion of "call backs." In performing a "One Word Story," human actors often make reference to actors, objects, locations, or actions that were previously mentioned in the story by another performer. To include this concept in our digital improvisers performance of a "One Word Story," the templates include places where the type based parameters are repeated, using "call backs" to give the story a cohesive feel. The "One Word Story," by the nature of its implementation, will also make call backs to the topics mentioned in the "Pattern Game."

Type Based Dictionaries

Given our set of twenty-five parameterized templates, we needed a way to compile sets of possible words that could fill the blanks, as indicated by their type. These types included categories such as female name, music genre, employee, emotion, school subject, action, fruit, appliance, etc.

To generate these lists, we rely on the hierarchical structure of WordNet. Given a type that is a noun, the objects in that list are comprised of the hyponyms of that noun. For example, the list of possible words for the type "athlete," is comprised of words that fall into the relationship "_____ is a kind of athlete" such as "runner" and "acrobat." Given a type that is a verb, the actions in that list are comprised of the troponyms of that verb. For example, the list of possible words for the type "walk," is comprised of words that fall into the relationship "_____ is a particular way to walk" such as "jog" or "waddle." Overall, the system uses 57 type based lists. The average length of a list is 127.8 words, with a median of 54. The longest list is "action words" which had 2,115 members while the shortest list, "sense words," has 5 members.

Selecting a Template

Since the "One Word Story" is a game which would typically follow a "Pattern Game" warm-up, it is natural for the actors to recall and tie the words from the "Pattern Game" context into the "One Word Story." That said, some contexts lend themselves more easily to particular templates. For example, the pattern game example mentioned earlier: "Music, fine art, art, creation, creative, inspiration, brainchild, product, production, magazine, newspaper, issue, exit, outlet, out" lends itself well to a template such as Story Template #3 above because Pattern Game context includes things that relate to artistic and musical productions/shows, topics which relate to the template and its parameter types.

Given a pattern game context, the system must choose an appropriate story template based on the context, as opposed to trying to make any template fit the topic matter. To do this, the system uses three sets of words, the words chosen for the Pattern Game, set *s*, the set of all words related to the words from the Pattern Game, set *s-rel*. Each template is represented by a set of words, set t_i, which is made up of the union of all of the type based lists for the parameter types used in that template. For each template, the system generates a score for how well the pattern game context matches the parameter types of the template. The score for the appropriateness of a template T_i, given a pattern game context, is calculated as the sum of the size of the intersection of set *s-rel* with set t_i and five times the size of the intersection set between set *s* and set t_i.

$$Appropriateness(T_i) = 5 * Count(t_i \cap s) + Count(t_i \cap s - rel)$$

Equation 6-1

From the scores that are calculated, the system chooses the highest as the most appropriate template, under the assumption that it relates most to the given pattern game context.

Given that some of the type based dictionaries are larger than others, one might assume that this calculation would make it more likely that some templates are chosen than others. While this is true, repetition of common templates did not emerge as an issue in the final system.

Selectional Restriction

Given a template and a Pattern Game context, the system must then find the most appropriate words to fill in this template to form a story. To make this judgment, the *Association Engine* uses a concept from the study of natural language processing called "selectional restriction." This is the concept that some combinations of words are more semantically valid than others. For example, it makes more sense to say "Joe ate the hamburger" than it does to say "Joe ate the butter" or even more so than "Joe ate the suspension" though all are nouns (hamburger, butter and suspension) and hamburger and butter are even both food products. In this system, I took a non-standard approach to selectional restriction. While I want to test if two words fit together, I don't use the standard rule base for doing so. I have found that such relations can be realized by mining one of the largest corpora of written language, the internet.

I use the internet to find the familiarity (popularity) of a particular verb/object pair. Given the nature of search engines, one can search by groupings of words or phrases. This has the side effect of allowing us to test whether certain groupings of words are used in documents that are indexed on the Web and to quantify their presence. Taking this approach to selectional restriction, we cannot generate new utterances, but can test whether certain utterances exist or how common they are. Using this information, we can elect phrases from which to form stories.

To apply selectional restriction in this case, I chose to again use the Web as a measure of common usage. For each blank in the template, the system generates a set of candidate fillers

based on the words from the Pattern Game and the words from the type based lists. For each of these candidates, a three word phrase, the candidate with the two words before and after the blank (not crossing punctuation), is queried against a search engine. The number of results for each query is used as a score for how commonly this phrase is used on the internet. This method serves us well as a way to form stories that sound natural.

For an example of how well this works, consider the sentence "The weather was _____." Assuming that we want to fill the blank with some "weather condition," let's consider the following candidates: stormy, rainy, rain, sunny, sun, awkward, and purple. The chart below shows the Google document frequencies from 3/20/07 of the candidate phrase "weather was ____" for all candidates. It can be seen that 'sunny' and 'rainy' would be the top choices based on this method, while 'purple' would be the least likely to be chosen.

Phrase	Google Document Frequency
"weather was stormy"	991
"weather was rainy"	19,900
"weather was rain"	1,930
"weather was purple"	0
"weather was sunny"	58,800
"weather was sun"	21
"weather was awkward"	7

Table 6-2 A set of phrases and their frequency of occurrence on the web.

The One Word Story Generated By Digital Improvisers

The final backend system is one that uses all the methods described above in order to generate a "One Word Story." The system begins with the set of words generated in a "Pattern Game" performance. From there, it chooses the most appropriate template using the method

described above. To fill the template, it chooses the words from the type based list that are most closely matched to the set of words from the pattern game, often using those exact words. When faced with a decision between words to choose, it uses selectional restriction to find the most appropriate word to fill a blank.

Below is a table containing two examples of "Pattern Games" and "One Word Stories," both generated by the *Association Engine*:

Pattern Game	One Word Story
"Music, fine art, art, creation, creative, inspiration, brainchild, product, production, magazine, newspaper, issue, exit, outlet, out."	"An artist named Colleen called her friend Alicia. Colleen wanted to go to the production at the music hall. Colleen and Alicia met up at the music hall. To their surprise there was no production at the music hall. Instead the women decided to go to the stage."
"Student, neophyte, initiate, tyro, freshman, starter, recruit, fledgling, entrant, beginner, novice, learner, apprentice, expert, commentator."	"There once was a woman named Lauren. Lauren wanted very much to be a student, but no one thought she could do it. To become a student, Lauren went to the institution, where all of the student people gather. Unfortunately, when Lauren got to the institution, she found out that all of the student people had become scholar people. Lauren felt diffidence."

Table 6-3 Two sample runs of the *Association Engine.*

PRESENTATION

The presentation of *The Association Engine* evolved greatly from its initial form. The first version of the system was one in which a viewer interacted with the system, using one screen to represent the digital improviser visually with a cloud of moving words, and using voice

recognition to take contributions from a human. The system would perform the "Pattern Game," creating a unified context between the machine and the human interacting. The next iteration of the system involved five screens that each displayed a cloud of moving words, representing five digital improvisers, where the role of the human was just to provide a seed word for the "Pattern Game." The final system embodied the actors with five animated faces, tied to a text to speech engine, who would perform the "Pattern Game" given a seed word from the audience, and finally produce a "One Word Story."

Interactive Model

The first version of the *Association Engine* consisted of a single digital improviser which would interact with a viewer. The digital improviser was represented by a single screen which displayed a pool of moving words. The viewer interacted with the system through a microphone, where contributions were interpreted by the system via voice recognition software.

The system began by prompting the user to say a word. After the user says a word, the system then takes that word and produces a word cloud on its display, similar to the could shown below where "Life" was the seed word, showing the space of words associated with the seed word. The words in the cloud move around for a few moments while the system acts as an improviser would, considering and internalizing the words in this association space. The seed word then shrinks to the size of the other words in the cloud. One of the other words begins to emerge and grow from the cloud. This is the word that the system chose to contribute to the "Pattern Game." The user then considers this word and makes another contribution related to this word and perhaps related to their original word as well. This process continues until the user is satisfied.

At its conception, this system was viewed as an interactive art installation, where the viewer could connect with a "digital improviser" and create a common context.

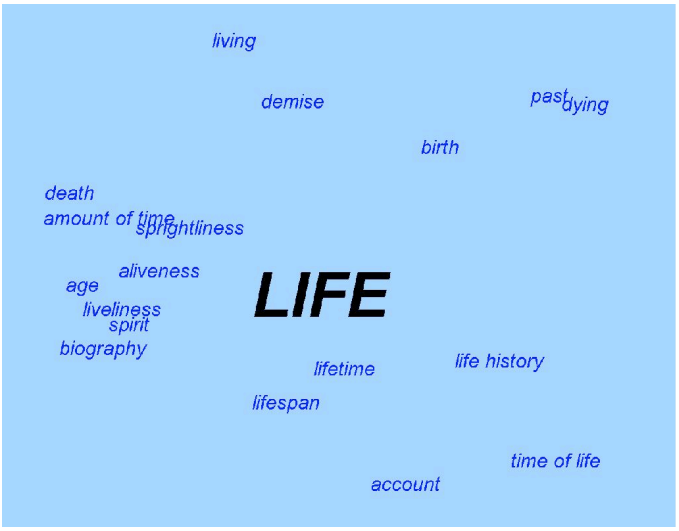


Figure 6-3 The first interface to the *Digital Improviser*.

A Team of Improvisers

While we found the interactive model of the *Association Engine* to be compelling, we also saw value in creating a free standing installation that scaled the idea of creating a digital improviser to the notion of a troupe of digital improvisers. Taking this scaling quite literally, we expanded the system to include five screens that each represented a digital improviser. Instead of using voice recognition for user input, we instead used keyboard input to take an initial seed from the user. In this system, I began to add other attributes to the words represented in the cloud space, conveying the types of the associations to the current word and the pattern game as a whole. The biggest change was that words are colored based on their relation to the current

word in focus (e.g. “synonyms” in red, “antonyms” in blue, “rhymes with” in green). We felt that adding some meaningful attributes to the words, enabled a richer experience.

Beyond an interactive art piece, we found this system to be useful as a brainstorming tool. It could become a tool for writers and artists looking for connections between disparate ideas. Brainstorming has become an important practice in group problem solving. It’s a way for group members to derive new knowledge from their collective previous knowledge. Each individual member comes to the table with his own background, vocabulary, and personal tendencies. Things that may be obvious to one participant might be new to other members of the group. By combining ideas, group members build associations and create a shared context from which to work.

Brainstorming works because each participant leaves the session with a much richer understanding of the problem domain. Using the *Association Engine*, a single individual could sit at a terminal and reap the benefits of a diverse collection of resources. Each of these resources could act like another participant in the brainstorming session, continually contributing ideas and drawing associations from other ideas offered. Since the sources could be varied in nature, a user would get the same richness of diversity experienced in a face-to-face meeting. The *Association Engine,* in this form, emulates and enhances the brainstorming process.

Five Actors

As an interactive art installation and a brainstorming tool, we found the *Association Engine* to be compelling, but as we began to extend beyond the “Pattern Game,” we chose to embody the digital improvisers. During the “Pattern Game,” a word based portrayal seemed appropriate, but as we started to move into the “One Word Story” as a performance, and attempting to give

the improvisers their own background knowledge and personalities, we found that the believability and connection to an improviser couldn't exist without embodiment.

In order to achieve this, we employed Ken Perlin's Responsive Face Project (Perlin 1996; Perlin and Goldberg 1996), pictured below. With Perlin's permission, we altered the polygon coordinates and colors to form four new models, shown below.

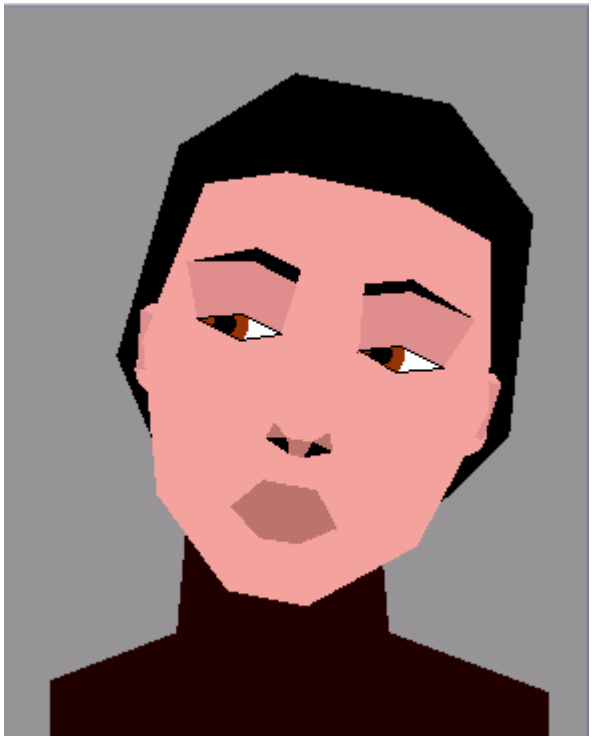


Figure 6-4 Ken Perlin's original Responsive Face.

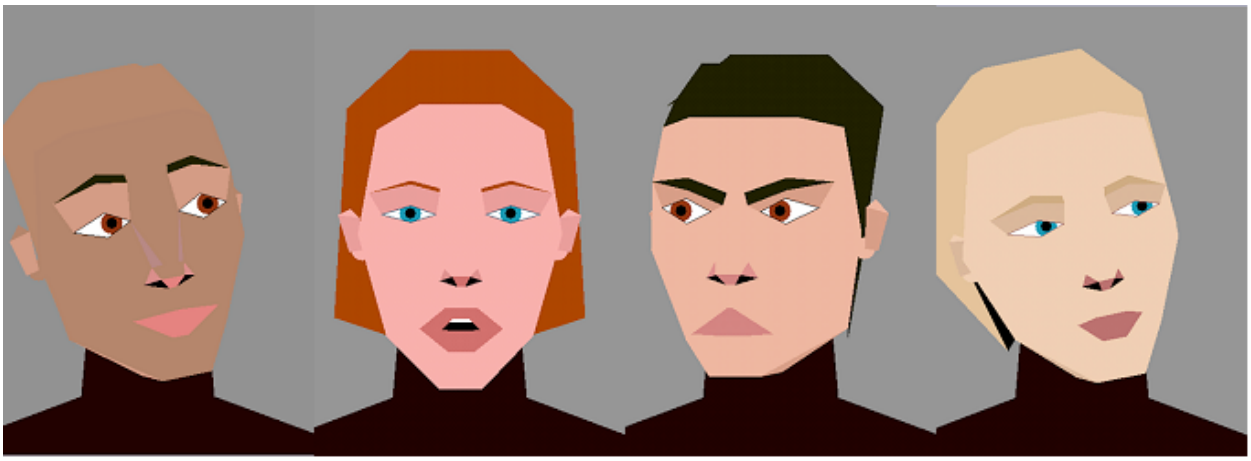


Figure 6-5 Four new faces, adapted from Perlin's original Responsive Face.

We connected the faces to a speech generation system, using a lip syncing approach. This approach will be discussed in more detail in future sections. The important part here is that we now have a team of five digital improvisers that are capable of voicing their contributions in an embodied avatar.

In addition to the five embodied digital actors, we chose to supplement the performance with a collective word space, representing the collection of words chosen so far, and the association space around them. For this, we used the same sort of word cloud that previously represented a single actor. To change the semantics, all chosen words are stored on the display in black, and all possible associations are given a different color to make the distinction. An image of the full installation including the five actors and supplemental display is shown in Figure 6-6.

Given our set of embodied actors, it soon became evident that using an embodied actor put more constraints on making the improvisers' interaction seem realistic. Things like timing, expression, turning, and tilting of the heads become much more meaningful. The digital actor must be empowered with some reasoning for how to act like a human improviser would.

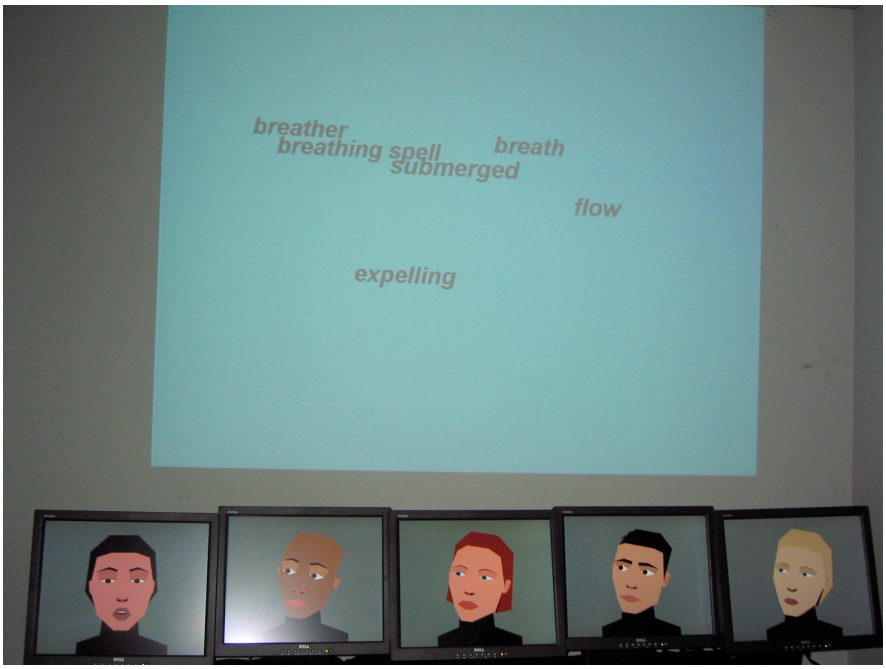


Figure 6-6 The *Association Engine* in action.

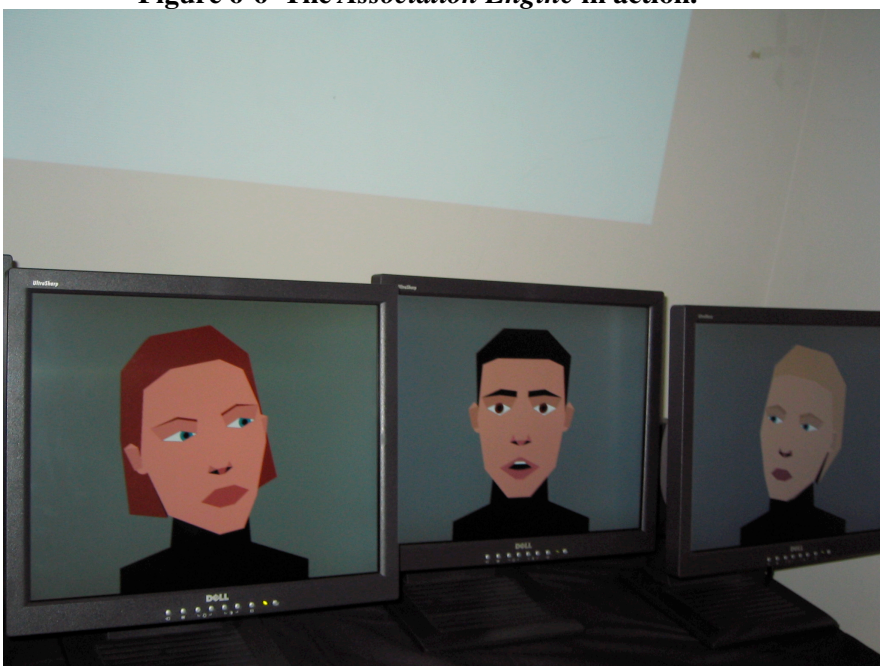


Figure 6-7 A close-up of the *Association Engine* actors.

In our first pass at building an embodied troupe of digital improvisers, we found that the associations were unrealistically quick, too fast for a human actor to possibly consider the

association space and make a contribution. Our first step at improving this involved a change made to the presentation of the "One Word Story." Instead of presenting just one word at a time, each actor would contribute a phrase to the story. The phrase consisted of one uncommon word (non stop word), and the remaining words before the uncommon word. For example, "up the hill" and "into the forest" are sample phrases. "Hill" and "forest" are uncommon words, while "up," "the," and "into" all occur on a list of common terms.

In addition to this change, we had to instill the actors with a notion of beats, that is, moments which have a meaning in the performance. The moments that became important or evident in the troupe were: listening to others, thinking of an association, and speaking an association. When listening to others, the digital actors are attentive by looking at the actor currently speaking, that is, their eyes and head are turned to face them, and their head is tilted to the side in a thoughtful position. While thinking of an association, it's important that the actors pause long enough to convey a genuine thought process. When speaking an association, the actor turns forward to face the audience.

INSTALLATIONS

The Association Engine was exhibited at the 2004 Performing Imagination Festival at Northwestern University. It was shown in its full form, as pictured in Figure 6-1. Viewers interacted with the system by typing in a seed word which began the "Pattern Game," followed by a "One Word Story." Feedback from the Festival was positive and grounded in a larger community including professors and students of film, studio art, theater, and performance art.

The *Association Engine's* "Pattern Game" backend was also exhibited as part of a larger installation in the 2004 PAC (Performing Arts Chicago) Edge Festival. The installation was

called *Handle with Care: Direct Mail and the American Dream*. The leading artist of the piece was GirlCharlie. The installation was set up as a staged living room with two recliners, a table, and walls covered in "direct mail." In the headrests of the recliners were installed sets of speakers. The speakers played the sounds of the *Association Engine*, computer generated voices, doing free association from seed words such as 'urgency," "danger," "accusation," "patriots," "outraged," "sick," "homosexuals," "vicious," and "imminent," all highly evocative and emotional words manually extracted from the real hate mail the lined the walls. The *Association Engine* was used to heighten the feelings of fear and fraud that one may feel when reading such mail. A full description of the piece can be found in Appendix 1.

CONCLUSION

While the *Association Engine* was not tremendously successful as a theater installation, the system not only brought our research into the space of story discovery (leading to *Buzz*), but also contributed to the success of *Buzz* in many ways. In developing the *Association Engine* we created controls for a set of actors, and developed a high level model for using these controls in a believable and realistic manner including lip syncing and semantic beats (described above). This model was carried over into the *Buzz* system, along with some of the textual classification methods used in the *Association Engine* including the familiarity metric. Most importantly, the *Association Engine* grabbed my interest in storytelling systems and criteria for evaluating compelling stories.

Chapter 7: Related Work

> *Knowledge, then, is experiences and stories, and intelligence is the apt use of experience and of the creation and telling of stories. Memory is memory for stories, and the major processes of memory are the creation, storage, and retrieval of stories.*
>
> - Roger C. Schank (Schank 1999)

Looking at the stories that the *Association Engine* produces, it is clear that the system faces problems that prevail from previous years of Artificial Intelligence research in story generation. The problems that these systems face, in accordance to their goals/motivations, are generating stories that are coherent, compelling and original. Previous story generation engines were able to generate stories that were coherent, based on a knowledge representation (engineering) approach, but because it was based on this knowledge representation, it was not scalable, so the stories are not novel. The larger problem is that the stories are not compelling or interesting. In general, the systems weren't built on a model of the aesthetic elements of a good story. How can we have a machine generate stories that are compelling if we haven't pinpointed or defined what makes a story compelling? These systems are an attempt to teach the machine to do something that not many humans are good at doing.

WHY STORY GENERATION?

Since much of Artificial Intelligence is concerned with understanding human intelligence, many researchers study how knowledge is acquired, represented and stored in our minds. Some theories of knowledge representation cite stories, also called cases, as the core representation schema in our minds. (Schank 1999) They explain that these stories are indexed in our memory by their attributes (locations, situations, etc.). As the quote at the beginning of this chapter explains, since stories are thought to be how memories represented, then the ability to understand, learn from, and tell stories could be seen as a measure of intelligence. From that metric, it's clear why many Artificial Intelligence researchers have focused their careers on building machines that are able to both understand and tell stories.

EARLY STORY GENERATION SYSTEMS

In the 1970s, many researchers began to build story generation systems, the most famous of which was *Tale-Spin*. *Tale-Spin* (Meehan 1977; Schank and Abelson 1977; Meehan 1981) used a world simulation model and planning approach for story generation. To generate stories, *Tale-Spin* triggered one of the characters with a goal and used natural language generation to narrate the plan for reaching that goal. The stories were simplistic in their content (using a limited amount of encoded knowledge) as well as their natural language generation.

Klein's *Automatic Novel Writer* (Klein, Aeschlimann et al. 1973) uses a world simulation model in order to produce murder novels from a set of locations, characters, and motivating personality qualities such as "dishonest" or "flirty." The stories follow a planning system's output as the characters searched for clues. The system does not seem to capture the qualities of a good murder story, including plot twists, foreshadowing, and erroneous clues.

Dehn's *Author* system (Dehn 1981) was driven by the need for an "author view" to tell stories, as opposed to the "character view" found in world simulation models. Dehn's explanation was that "the author's purpose is to make up a good story" whereas the "character's purpose is to have things go well for himself and for those he cares about"(Dehn 1981).

These three systems are a good representation of the early research done in story generation. They cover both approaches that are structuralist, in the case of the *Automatic Novel Writer* where a predefined story structure is encoded, and transformationalist, in the cases of *Tale-Spin* and *Author* (Liu and Singh 2002) which generated stories based on a set of rules or goals.

RECENT APPROACHES TO STORY GENERATION

Over the past several decades, research in story generation has continued. Recent work has taken a new spin; using various modern approaches in an attempt to solve this classic AI problem. (Smith and Bates 1989; Murray 1997; Mateas and Sengers 1999) *Make-Believe* (Liu and Singh 2002) makes use of the structure and content of Open Mind Commonsense (OMCS), a large-scale knowledge base, in order to form stories. It extracts cause-effect relationships from OMCS and represents them in an accessible manner. After taking a sentence from the user to start the story, it forms a causal chain starting with the state represented in the sentence. The end result is a simple story that represents a chain of events. Similar to *Make-Believe*, many others have taken the approach of interactive story telling systems, which take advantage of the creativity of their users in providing seed sentences and on going interactions (Mateas, Domike et al. 1999; Mateas 2001; Domike, Mateas et al. 2003).

Brutus (Bringsjord and Ferrucci 2000; Sousa 2000) uses a more knowledge representation intensive method to tell stories with literary themes, such as betrayal. The stories often

intentionally omit some key details to leave the reader with a bit of mystery. The stories are more interesting, but at the cost of being less automated since each portrayal of an interesting theme must be represented for the system.

Some recent systems have taken a Case Based Reasoning approach to story and plot generation (Gervas, Dıaz-Agudo et al. 2005). *Minstrel* (Oatley 1994; Turner 1994) uses case based reasoning, still within the confines of a problem-solving approach to story generation. The case base is a hand coded based of known stories, which is used to extract plot structures. The goals for the problem solving system are devised from an author view (as Dehn proposed) as opposed to a character point of view. These goals exist across four categories: dramatic, presentation, consistency and thematic. This system, along with *Brutus*, seem to be a huge step closer to generating stories that are interesting to a reader and that embody many of the traits of good stories.

STORY DISCOVERY

In general, previous story generation systems faced a trade-off between a scalable system, and one that can generate coherent stories. Besides Dehn's *Author*, early research in this area employed a weak model of the aesthetic constraints of story telling. More recent work on story generation has taken a more interactive approach, leverage input from humans, or taking a case based approach to achieve greater variance.

In response to the shortcomings of story generation, and taking the notion of leveraging input from humans a step further, I chose to explore story discovery. The emergence of weblogs (blogs) and other types of user generated content has transformed the internet from a professionally produced collection of sites into a vast compilation of personal thoughts,

experiences, opinions and feelings. The movement of the production of online news, stories, videos and images beyond the desks of major production companies to personal computers has provided internet users with access to the feelings and experiences of many as opposed to merely professionally edited content.

While this collection of blogs is vast, only a subset of blog entries contains subjectively interesting stories. Using what I learned from previous story generation systems to inform story discovery, I define a stronger model for the aesthetic elements of stories and use that to drive retrieval of stories, and to filter and evaluate results. *Buzz*, described in Chapter 2, is a system which exemplifies this notion of story discovery.

RELATED ARTISTIC SYSTEMS

Artistically, story telling and online communication have been externalized within several installations. Of the more well known in this area, *Listening Post* (Hansen and Rubin 2002; Hansen and Rubin 2004) is an art installation that exposes content from thousands of chat rooms and other online forums through an audio and visual display. The audio component is made up of synthesized voice of the content which is displayed visually on more than two hundred tiny screens. *Listening Post* exposes the real-time 'buzz' of chat rooms on the web, in a manner that was designed to convey "the magnitude and diversity of online communication."

Inspired by *Listening Post*, We *Feel Fine* (Harris and Kamvar 2007) is an ongoing online art installation that was created in August of 2005 by Jonathan Harris and Sep Kamvar. The backend for the installation mines the blogosphere for sentences containing the phrase "I feel" or "I am feeling." It discovers 15 to 20 thousand new sentences per day and for each sentence, when possible, extracts demographic information (age, gender, location, etc.) and any image

associated with the blog entry. Using the time of the post and the location information, it gathers information about the current weather at the blogger's location. It also does a simple association of the sentence to zero or more of five thousand hand coded emotions. The associations are determined by mere presence of the emotion term in the sentence. The system has mined and stored these sentences since August of 2005 and visualized them in several interesting ways through a web applet.

Like both *Listening Post* and *We Feel Fine*, *Buzz* externalizes online communication, through a context refined via search and extraction. The *Buzz* filters and retrieval strategies focus the presentation on stories and specifically on compelling and emotional stories while these two installations are more broadly oriented to represent or portray the breadth of online communication, while also targeting key phrases. *Listening Post* demonstrates online communication as a whole, while *Buzz* focuses on singular voices from the blogosphere, grounded in current popular topics and/or dramatic situations. *Buzz* also differs in that the bloggers are embodied by avatars to give voice and body to the stories.

Chapter 8: Conclusions and the Future

71 million blogs…some of them have *to be good.*

- Matt (Technorati 2007)

In this dissertation, I've described a set of systems that connect people; through stories, opinions, and points of view. These stories are automatically extracted from blogs, and presented in a range of systems. People tell stories to connect to other people. These systems amplify the stories that people tell and facilitate and strengthen connections between people.

The first of these systems, *Buzz*, emerged as a digital theater installation from an area called Network Arts. For this system, I've enabled a team of virtual actors with a toolset to find stories in the blogosphere, measure their impact, and present the most meaningful or compelling stories to an audience in an engaging way. Not only is this enjoyable from the audience's perspective, but the system amplifies many previously unheard voices in the blogosphere by presenting them in a public installation where thousands of people will hear them.

Buzz also exists in a mode to connect companies to their consumers, through consumer's stories of experiences with and opinions about brands and products from the blogosphere. In this mode, the team of virtual actors uses the same toolset to find stories, measure their impact, and present the most interesting stories. However, the metrics for interestingness differ from the base

system, as the audience and goals of the system have changed. While the system still functions to connect people through stories, the stories that it presents must be completely relevant to the brands and products that particular companies care about. It also follows that measures of emotion must be specific to the product or brand mentioned, not the general sentiment of the story.

Finally, *News at Seven* is a system which makes use of the *Buzz* story discovery engine in order to find emotional points of view about a news story, and present them to an audience. As in the brand and product based *Buzz*, relevance is crucial to the impact of the points of view presented in *News at Seven*. The system is configured to look for stories that sound like opinions as opposed to stories that take a general narrative form. For *News at Seven,* the *Buzz* story discovery engine performs well as a way of automatically mining opinions about current events, and as a result achieves the goal of connecting people through opinions and points of view.

In addition to autonomously finding stories and judging interestingness, a concept that extends to search by interestingness, the *Buzz* story discovery engine has broader implications and contributions. Consider how people tell stories – they never "create" or "invent" new stories, instead they recall the most interesting stories they've heard or experiences they've had, they may merge different experiences together, and they alter the details to make a more compelling story. In a sense, we've codified how people tell or create stories, building a system to do just that. In a way, I think these systems made contributions to the decades old problem of story generation in Artificial Intelligence.

The clear next step for *Buzz* includes moving it online, where instead of connecting thousands of people in a public installation, it could impact millions. Future work in this area involves creating a destination entertainment site (www.buzz.com), where viewers can see

interesting stories presented to them by a diverse set of realistic and engaging avatars. The stories will exist in a variety of high level categories, covering the types of stories presented in all three systems described above. This includes compelling stories involving dramatic situations (dreams, fights, confessions, etc.), topical stories (around new products or hot topics), and news related points of view (evolving daily). Each day, Buzz.com will produce and host approximately 1400 new 45 to 90 second video stories based on the updated daily news events, topics, etc. The videos can be navigated via topical search or browsed through a set of hierarchical categories. The site will allow users to comment on videos, rate them, and recommend them to friends. We hope that *Buzz.com* will be successful in connection millions of people through stories of experiences and points of view.

STORY DISCOVERY

I've provided these systems with a toolset to find stories, measure their impact, and present the most interesting stories. To find stories and measure their impact, these systems use a retrieval, filtering and modification model that takes advantage of the vastness of the blogosphere by aggressively filtering the retrieval of stories. The filters act as an editor of a magazine or journal, selecting the most interesting and appropriate content, while the modifiers amplify this content. Three types of filters and modifiers have been crucial to this toolset: *Story Filters* which narrow the blogosphere down to a set of stories and *Story Modifiers* which modify the candidate stories to a more story like structure, *Content* or *Impact Filters* which further narrow the set to those that are appropriate and impactful and *Content or Impact Modifiers* which amplify the content of the candidate stories, *Presentation Filters* which evaluate how compelling

each story will be when presented by a computer generated voice and avatar and *Presentation Modifiers* which alter the story candidates to make them sound more appropriate when presented.

As we move forward with Buzz.com, additional filters and possibly even additional filter types will need to be generated, as driven by the need of this large scale system with new categories of stories. One of the strengths of this model is that the filters are fully configurable to meet the need of a new system. Furthermore, the model is additive; additional needs that arise can be handled by the addition of new filters to the set.

The new system for *Buzz.com* must also be enabled with a metric to evaluate the interestingness of topics and current events as they evolve in the world and drive story discovery accordingly. Allowing Buzz.com to automatically discover and evaluate topics is crucial to making this site interesting to the user, while remaining autonomous.

STORY PRESENTATION

In presenting stories, we found it compelling to have them told by an animated avatar tied to voice generation software. Embodying the blogger is quite powerful in allowing the viewer to situate themselves in the narrative. To back this notion, we've found that many researchers in Artificial Intelligence have said that both hearing concepts presented as stories, and seeing/hearing these stories presented in an interesting way, is the most powerful way to learn or take in information (Schank 1990).

As *Buzz.com* moves forward, the site must use more realistic looking avatars. The avatars will be 3D rendered models of a body from the shoulders up. Each category will have a set of appropriate representative avatars to present the stories for that category. To make the

presentation compelling, the avatars must be driven by a stronger model of emotional speech and gesture. To do this, we may extend the emotional speech emphasis model created for *Buzz* to provide more variation in the generated speech. This model may also be extended to allow for variance in the speech patterns for the different avatars. While it is important for the avatars to look realistic and engage the audience, it is equally critical that the avatars appearance and animations do not exceed the quality of their voice. Such disconnect, between voice and animation quality will prove to be distracting for the viewers (Mori 1970).

STORIES

Living in a world where the machine and the Internet are ubiquitous, many people work and play online, in a world that is, ironically, often isolated and lonesome. While the Internet, as intended, connects us to information, products and services, it often draws us away from the rich connections that are created through interpersonal communication. People tell stories as a way to be less lonesome, reaching out to people that they can relate to. We have ready and willing participants, millions of bloggers who are reaching out to people by telling stories. The systems that I've described in this dissertation are intended to facilitate and amplify these connections, using the very machine that pulls us apart, to bring us together, connecting people through stories.

> *We are lonesome animals. We spend all our life trying to be less lonesome. One of our ancient methods is to tell a story begging the listener to say -- and to feel -- Yes, that's the way it is, or at least that's the way I feel it. You're not as alone as you thought.*
>
> - John Steinbeck (Steinbeck 1954)

Appendix

1. HANDLE WITH CARE: DIRECT MAIL AND THE AMERICAN DREAM BY GIRLCHARLIE

Below is a full description of this installation from the 2004 Chicago PAC (Performing Arts Chicago) Edge festival program.

"Direct mail inundates us with requests for and offers of money. It promises us untold riches if we only buy this one thing and untold danger if we don't. This installation focuses on the particularly manipulative invective of mail from sweepstakes and right-wing political causes and targets older and new Americans in devious ways, playing on hope and creating fear and hate.

Direct mail is niche marketing. It whispers particular words to particular people. It isn't meant to be broadcast to the world. This installation is dedicated to pulling it out of the shadows. To making visible the nature and the volume of direct mail.

It wheedles, insinuates, accuses, and demands money from people by fomenting fear, anxiety and suspicion. It presents false hopes and false gods all to draw more dollars from people's pockets. It is insanely personal. It looks like official government mail. It looks like official government checks. It looks like a personal plea from a personal friend – someone who knows you in your heart. And all you have to do is reply. Send in a check. Engage.

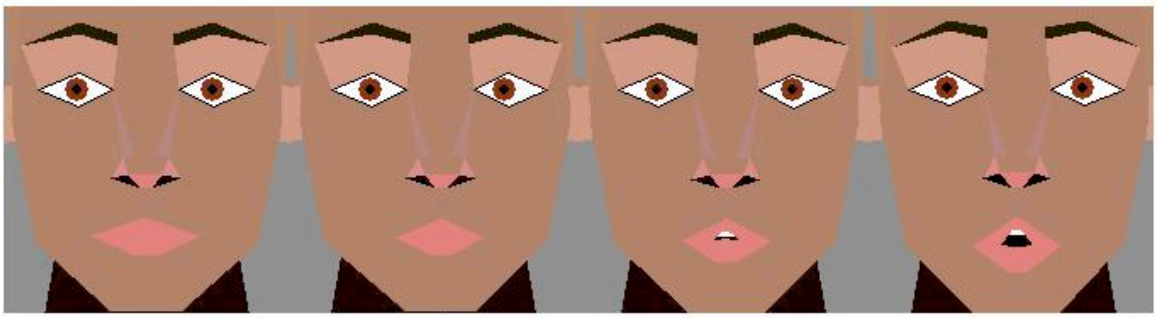
Unsolicited mail exploits our most vulnerable communities, in particular elderly Americans and new Americans who are used to a world where every piece of mail is a sincere missive.

Often they don't have the experience yet to identify this particularly insidious form as fraud, to know how to filter it out, to resist the urge to read it and fall victim to its ostensible authenticity."

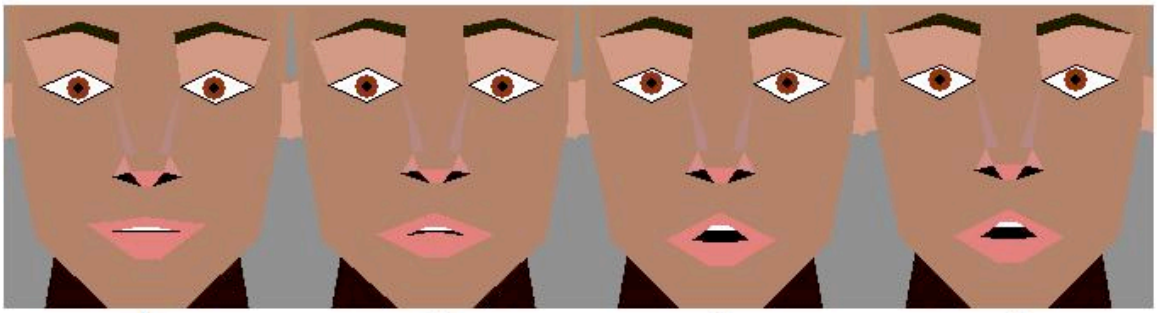
2. MSAPI VISEME TO IMAGE MAPPING

MSAPI Viseme #	Phoneme	Lip Position Image #
0	Silence	0
1	AE, AX, AH	11
2	AA	11
3	AO	11
4	EY, EH, UH	10
5	ER	11
6	Y, IY, IH, IX	9
7	W, UW	2
8	OW	13
9	AW	9
10	OY	12
11	AY	11
12	H	9
13	R	3
14	L	6
15	s, z	7
16	SH, CH, JH, ZH	8
17	TH, DH	5
18	f, v	4
19	d, t, n	7
20	k, g, NG	9
21	p, b, m	1

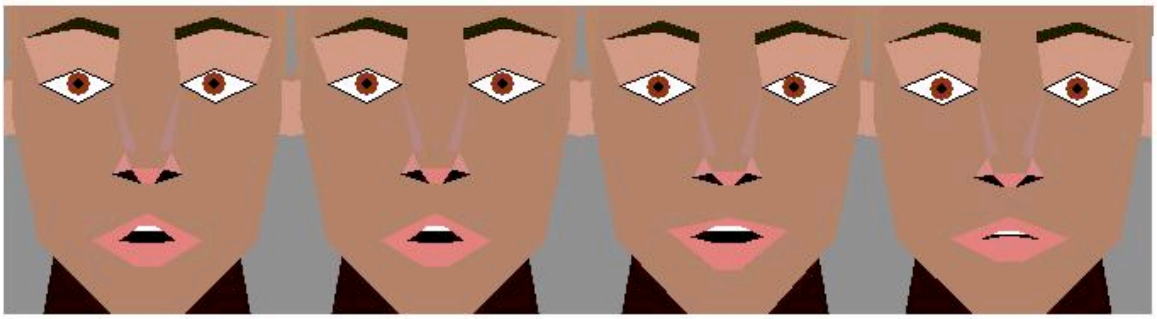
Lip Position Images



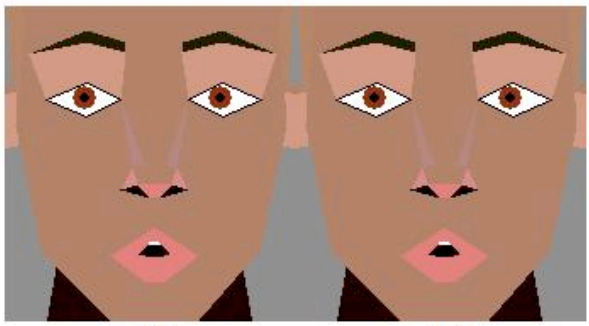
0 1 2 3



4 5 6 7



8 9 10 11



12 13

3. LEVIN VERB CLASS 31.2: ADMIRE VERBS (LEVIN 1993)

- treasure
- support
- lament
- hate
- fancy
- respect
- resent
- pity
- loathe
- abhor
- detest
- favor
- prize
- idolize
- distrust
- revere
- enjoy
- stand
- value
- like
- disdain
- deplore
- trust
- admire
- fear
- venerate
- appreciate
- rue
- exalt
- dislike
- adore
- regret
- envy
- execrate
- tolerate
- mourn
- despise
- savor
- relish
- miss
- cherish

- esteem
- worship
- love
- dread

References

AggressiveResponse. (2007). "http://aggressiveresponse.wordpress.com/2007/06/12/this-week-in-palestine/."

Apache (2007). Apache Lucene.

Aue, A. and M. Gamon (2005). Customizing sentiment classifiers to new domains: a case study. RANLP.

Beeferman, D. (1998). Lexical discovery with an enriched semantic network. Workshop on Applications of WordNet in Natural Language Processing Systems, ACL/COLING.

Bekkerman, R. and J. Allan (2004). Using Bigrams in Text Categorization. Amherst, MA, University of Massachusetts.

Benamara, F., C. Cesarano, et al. (2007). Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone. International Conference on Weblogs and Social Media. Boulder, Colorado.

Black, A. (2002). Perfect synthesis for all of the people all of the time. IEEE TTS Workshop. Santa Monica, CA.

Bringsjord, S. and D. A. Ferrucci (2000). Artificial Intelligence and Literary Creativity. Mahwah, NJ, Erlbaum.

Bruninghaus, S. and K. D. Ashley (1998). How machine learning can be beneficial for textual case-based reasoning. AAAI-98/ICML-98 Workshop on Learning for Text Categorization.

BusinessWeek. (2007). from http://www.businessweek.com/the_thread/blogspotting/archives/2007/04/blogging_growth.html.

Cahn, J. E. (1990). "The Generation of Affect in Synthesized Speech." Journal of the American Voice I/O Society.

Chesley, P., B. Vincent, et al. (2006). Using Verbs and Adjectives to Automatically Classify Blog Sentiment. AAAI Spring Symposium "Computational Approaches to Analyzing Weblogs". Stanford University.

Csikszentmihalyi, M. (1991). Flow: The Psychology of Optimal Experience, Harper Perennial.

Cunningham, P., N. Nowlan, et al. (2003). A case-based approach to spam filtering that can track concept drift. ICCBR Workshop on Long-Lived CBR Systems.

Cymfony. (2007). "Cymfony." from http://www.cymfony.com/.

Dehn, N. (1981). Story Generation after Tale-Spin. Seventh Internation Joint Conference on Artificial Intelligence. University of British Columbia.

Domike, S., M. Mateas, et al., Eds. (2003). The recombinant history apparatus presents: Terminal Time. Narrative Intelligence. Amsterdam, John Benjamins.

Ekman, P. (2003). Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. New York, NY, Henry Holt and Company.

Finn, A. and N. Kushmerick (2003). Learning to classify documents according to genre. IJCAI-03 Workshop on Computational Approaches to Style Analysis and Synthesis.

Gervas, P., B. Dıaz-Agudo, et al. (2005). "Story plot generation based on CBR." Knowledge based systems **18**(4-5): 235-242.

Google. (1996). "Google." from http://www.google.com.

Hansen, M. and B. Rubin (2002). Listening Post: Giving Voice to Online Communications. International Conference on Auditory Display. Kyoto, Japan.

Hansen, M. and B. Rubin. (2004). "Listening Post." from http://earstudio.com/projects/listeningpost.html.

Harris, J. and S. Kamvar. (2007). "We Feel Fine." from http://www.wefeelfine.org/.

Healy, M., S. Delany, et al. (2005). An assessment of case-based reasoning for short text message classification. Sixteenth Irish Conference on Artificial Intelligence and Cognitive Science.

Hoffmann, R. (2000). "Narrative." American Scientist Online **88**(4).

Kamps, J., M. Marx, et al. (2004). Using WordNet to measure semantic orientations of adjectives. Fourth International Conference on Language Resources and Evaluation.

Klein, S., J. Aeschlimann, et al. (1973). Automatic novel writing., University of Wisconsin Madison.

Koppel, M., S. Argamon, et al. (2003). "Automatically categorizing written texts by author gender." Literary and Linguistic Computing.

Levin, B. (1993). English Verb Classes And Alternations: A Preliminary Investigation. Chicago, The University of Chicago Press.

Liu, H., H. Lieberman, et al. (2003). A model of textual affect sensing using real-world knowledge. Intelligent User Interfaces, ACM Press**:** 125-132.

Liu, H. and P. Singh (2002). MakeBelieve: Using commonsense knowledge to generate stories. Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence.

Mateas, M. (2001). "Expressive AI: A hybrid art and science practice." Leonardo: Journal of the International Society for Arts, Sciences, and Technology **34**(2): 147-153.

Mateas, M., S. Domike, et al. (1999). Terminal time: An ideologically-biased history machine. AISB Symposium on Artificial Intelligence and Creative Language: Stories and Humor.

Mateas, M. and P. Sengers (1999). Narrative Intelligence. AAAI 1999 Fall Symposium Series.

Meehan, J. R. (1977). Tale-spin, an interactive program that writes stories. the 5th IJCAI.

Meehan, J. R. (1981). TALE-SPIN and Micro TALE-SPIN. Inside Computer Understanding. R. Schank and C. Riesbeck. Hillsdale, NJ, Erlbaum**:** 197 to 258.

Mori, M. (1970). "The Uncanny Valley." Energy **7**(4): 33-35.

Muresan, S. (2001). Combining Linguistic and Machine Learning Techniques for Email. Annual Meeting of the ACL, Workshop on Computational Natural Language Learning.

Murray, J. (1997). Hamlet on the Holodeck: The Future of Narrative in Cyberspace, The Free Press.

Napier, M. (2004). Improvise: Scene From The Inside Out. Portsmouth, NH, Heinemann.

NeoSpeech (2006). NeoSpeech.

Nichols, N., K. Hammond, et al. (2006). Believable Performance Agents for Interactive Conversations. ACM SIGCHI Advances in Computer Entertainment. Hollywood, California, ACM Press.

Nichols, N., S. Owsley, et al. (2007). News at Seven: An Automatically Generated News Show. submitted to ACM MultiMedia.

NielsenBuzzMetrics. (2007). "NielsenBuzzMetrics." from http://www.nielsenbuzzmetrics.com/.

Oatley, K. (1994). "The Creative Process: A computer model of storytelling and creativity." Computational Linguistics **21**(4): 579-581.

Ortony, A., G. L. Clore, et al. (1987). "The referential structure of the affective lexicon." Cognitive Science **11**: 341-362.

Owsley, S., S. Sood, et al. (2006). Domain Specific Affective Classification of Documents. AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs. Palo Alto, California.

Perlin, K. (1996). "Responsive Face Project."

Perlin, K. and A. Goldberg (1996). "Improv: A System for Scripting Interactive Actors in Virtual Worlds." Computer Graphics **29**(3).

Pinkney, J. (2000). Aesop's Fables, SeaStar.

Polti, G. and L. Ray (1940). The Thirty-Six Dramatic Situations. Boston, Writer.

Porter, M. F. (1980). "An Algorithm for Suffix Stripping." Program **14**(3): 130-137.

Price, R. and L. Stern (1974). The Original Mad Libs 1, Price Stern Sloan.

RateItAll. (2006). "Rate It All: The Opinion Network." from www.rateitall.com.

Raux, A. and A. Black (2003). A unit selection approach to f0 modeling and its application to emphasis. ASRU. St. Thomas, US Virgin Islands.

RomanceBlogReader. (2007). "Intellect vs Emotion." from http://romblogreader.blogspot.com/2007/01/intellect-vs-emotion.html.

Salton, G. (1983). Introduction to Modern Information Retrieval, McGraw-Hill.

Salton, G. and C. Buckley (1988). "Term-weighting approaches in automatic text retrieval." Information Processing and Management. **24**(5): 513 to 523.

Schank, R. C. (1990). Tell Me A Story. Evanston, IL, Northwestern University Press.

Schank, R. C. (1999). Dynamic Memory Revisited. Cambridge, United Kingdom, Cambridge University Press.

Schank, R. C. and R. Abelson (1977). Scripts Plans Goals and Understanding: An Inquiry into Human Knowledge Structures. Hillsdale, New Jersey, Lawrence Erlbaum Associates, Publishers.

Shamma, D., S. Owsley, et al. (2004). Using Web Frequency Within Multimedia Exhibitions. ACM Multimedia. New York, ACM Press.

Shamma, D. A. (2005). Network Arts: Defining Emotional Interaction in Media Arts and Information Retrieval. Computer Science Department. Evanston, IL, Northwestern University. **Doctor of Philosophy**.

Shamma, D. A., S. Owsley, et al. (2004). Network Arts: Exposing Cultural Reality. The International World Wide Web Conference. New York.

Simmons, A. (2002). The Story Factor: Inspiration, Influence, and Persuasion Through the Art of Storytelling, Perseus Books Group.

Smith, S. and J. Bates (1989). Towards a theory of narrative for interactive fiction. Pittsburgh, PA, Carnegie Melon University.

Sood, S., S. Owsley, et al. (2007). Reasoning Through Search: A Novel Approach to Sentiment Classification. submitted to International World Wide Web Conference.

Sousa, R. d. (2000). "Artificial Intelligence and Literacy Creativity: Inside the Mind of BRUTUS, a Storytelling Machine." Computational Linguistics **26**(4): 642-647.

Sproat, R., M. Ostendorf, et al. (1998). The Need for Increased Speech Synthesis Research: Report of the 1998 NSF Workshop for Discussing Research Priorities and Evaluation Strategies in Speech Synthesis, NSF.

Steinbeck, J. (1954). In Awe of Words, Exonian.

Technorati. (2007). "Technorati." from www.technorati.com.

Turner, S. R. (1994). The Creative Process: A computer model of storytelling and creativity. Hillsdale, NJ, Lawrence Erlbaum Associates.

Umbria. (2007). "Umbria." from http://www.umbrialistens.com/.

Wikipedia. (2005). "Wikipedia." from http://www.wikipedia.org/.

WordNet. (1997). "WordNet." 2004, from http://wordnet.princeton.edu/

Yahoo! (2006). "Yahoo! ." from http://www.yahoo.com.

Zemeckis, R. (2000). Cast Away.

Zipf, G. (1949). Human Behavior and the Principle of Least-effort. Cambridge, MA, Addison-Wesley.