The Automatist Storytelling System
Putting the Editor's Knowledge in Software

by
Michael Luke Murtaugh

B.S. Computer Science and Engineering
Massachusetts Institute of Technology
Cambridge, MA
1994

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MEDIA TECHNOLOGY
at the
Massachusetts Institute of Technology
September 1996

Signature of Author
Program in Media Arts and Sciences
August 9, 1996

Certified by
Glorianna Davenport
Principal Research Associate
Media Laboratory
Thesis Supervisor

Accepted by
Stephen A. Benton
Chairperson
Departmental Committee on Graduate Students
Program in Media Arts and Sciences

The Automatist Storytelling System
Putting the Editor's Knowledge in Software

by
Michael Luke Murtaugh

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, on August 9, 1996,
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MEDIA TECHNOLOGY
at the
Massachusetts Institute of Technology
September 1996

Abstract

This thesis presents the Automatist Storytelling System—an "editor in software" or "narrative engine"—a system that produces dynamic and responsive presentations from an extensible collection of keyword-annotated materials. Sequencing decisions are made on the basis of association, and the overall structure and meaning of an experience emerges from the interactions of individual material presentations. In this highly decentralized model, viewers are consistently integrated participants, who exert varying degrees of influence or control over the construction of the experience. The viewers' role is considered primarily extradiegetic; viewers' actions influence the process of the storytelling rather than altering actual events in the story world. By making both the viewing experience and authoring process variable and extensible, the Automatist Storytelling System supports new story forms such as the "Evolving Documentary."

This thesis presents two systems, ConTour and Dexter, as examples of Automatist Storytelling Systems. These systems were developed and are described in terms of, respectively, the stories: *Boston: Renewed Vistas* and *Jerome B. Wiesner: A Random Walk through the Twentieth Century.*

Thesis Supervisor: Glorianna Davenport
Title: Principal Research Associate

This work was supported in part by the Eastman Kodak Company.

The Automatist Storytelling System
Putting the Editor's Knowledge in Software

by
Michael Luke Murtaugh

The following people served as readers for this thesis:

Reader: _____

Kenneth W. Haase, Jr.
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Reader: _____

Toni Dove
Electronic Media Artist

Acknowledgments

Thanks first to my parents for their support.

Thanks to IC past and present: Ryan Evans, Eddie Elliott, Tinsley Galyean, Lee Morgenroth, David Tamés, David Kung, Kevin Brooks, Natalia Tsarkova, Freedom Baird, Richard Lachman, Brian Bradley, and Peter Cho. Thanks also to my teachers: Suguru Ishizaki, Mitchell Resnick, Janet Murray, Marvin Minsky, Ron MacNeil, and Larry Friedlander.

Special thanks to Linda Peterson and Santina Tonelli for their patience and support.

Particular thanks to Sawad Brooks, for both friendship and guidance, one result of which is a supremely better informed discussion of interactive narrative in this thesis.

Thanks to my readers Ken Haase and Toni Dove, whose work and teachings continue to inspire me.

Special thanks to those whose friendship continues to sustain me: Robin Kullberg, Dinko Zidaric, and Betsy Brown.

Bonus thanks to Robin and Peter for their invaluable assistance editing this text, and for their feedback and encouragement.

Finally this thesis stands as one end of a more than two year conversation with my advisor, Glorianna Davenport, whose unique approach and vision informs every page.

Table of Contents

Introduction

Interaction & Narrative

Before considering the relationship of interaction to narrative, we need to be clear about our terminology regarding narrative. Gérard Genette begins his essay *Narrative Discourse* with the following distinctions: (1) the *story* is the signified or narrative content described, (2) the *narrative* is the signifier or form of the narrative discourse itself, and (3) the *narrating* is the "producing act."[1] Thus, narrative in a general sense is presented as a system where the act of storytelling, the narrating, produces a narrative through which the viewer constructs a story world, or *diegesis*.

The introduction of interactivity complicates this picture. In the novel and film, the form of the discourse—the narrative—exists as a static object to be presented the same way to all viewers. With interactivity, the assumption is that the narrative varies. Interactivity could thus be said to occur at the level of the narrating. Each interactive reading or viewing represents a separate "producing act" the result of which is a narrative specific to that viewing. Thus when we speak of the *narrative* of an interactive narrative, we refer not to a static form but to the potential and inherently variable viewer *experience* of its construction.

The Constraints of Mass Audience

Interactivity raises the question: *who's telling the story?* While the ultimate understanding of any narrative depends on both the author and the viewer, authors have generally exerted explicit control over the exact form of the narrative.

[1] Genette, Gérard. *Narrative Discourse.* Cornell University Press, 1980. p. 3

Seen this way, a key question becomes: why should an author relinquish control over the narrative? Electronic media artist Toni Dove recalls her initial response to the idea of incorporating interaction in her work as "Why should I replace intellectual challenge with multiple choice?"[2] Conversely, one might ask: why should a viewer want to exert control over the narrative? While the idea of interactivity remains appealing to the general public, people aren't exactly restless in their theater seats for something to click on.

To begin, an author may be willing to release some control if in return such a loss provides additional benefits as prior constraints are removed.

For example, documentary films must often conform to a rigid structure imposed by the conditions of their presentation. A program might need to be a specific length or be structured in a particular way to facilitate television scheduling. At the same time, filmmakers generally gather much more content than they can fit into their allotted time slots.

The inherent necessity to produce one specific "cut" of a story places a limitation on the use of content. Structuring a story around a particular theme might prevent the full incorporation of available material about another. Furthermore, presenting only a single structure limits a story from potentially taking on a variety of meanings through alternative tellings. The variability of the interactive narrative's form enables a multiplicity of story meaning.

Even without constraints imposed by the presentation, additional constraints relating to the audience exist. Lacking knowledge about each specific viewer, the author is instead asked to gauge the range of their audience and, in the interest of broadest possible appeal, structure the narrative to please the lowest common denominator.

For instance, lacking knowledge of the amount of time a specific viewer has available or is willing to spend, the author is obliged to conform to some "practical" time limit given their content.

[2] Dove, Toni. "Theater without Actors—Immersion and Response in Installation" *Leonardo*, V27, No. 4, 1994, p. 281

Lacking information about a specific viewer's interest or knowledge, the author is similarly asked to tailor story content to a vague notion of "broad interest" or appeal. The result is an absence of any significant depth on issues not believed to be of general interest.

In sum, lowest common denominator programming places depth of content in inverse relation to breadth of audience; the result is shallow and disjoint "sound bite" programming.

The converse of the above is that relinquishing certain aspects of authorial control enables viewers to form a more personal and meaningful connection to the story. In this way, interactivity may function to increase viewer engagement with the narrative by facilitating a specific viewer's knowledge and viewing situation. As a form that supports multiple meanings, the interactive narrative has the potential to tell more complex and personally meaningful stories than those delivered to a mass audience.

Hypermedia

Many approaches to incorporating interactivity with narrative have been based on a literal notion of making viewer actions map into story actions—placing the viewer in the role of a character. Consequently, interactive stories based on this model have been used to tell stories about making choices, exploring the results of the actions we choose to make. Such interactive stories presumably encourage the viewer to role-play and thus "learn by doing."

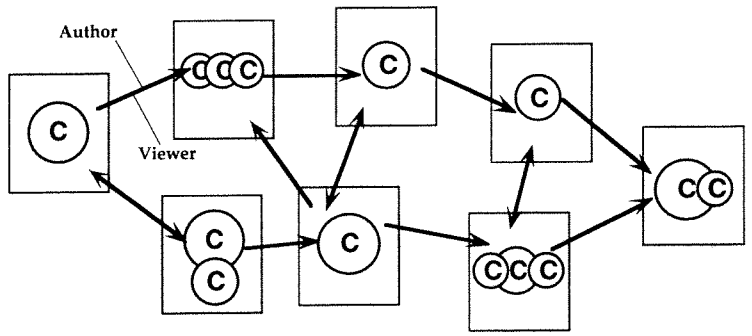
Unfortunately the experience of such stories often seems empty. Choices seem contrived and constrained and the experience seems flat and ultimately devoid of meaning. Toni Dove attributes this effect to a "fallacy of choice" and finds the process of making decisions within an author's preconceived "matrix" of possibilities analogous to the "false sense of choice" instilled by advertising. [3]

Furthermore, making a viewer's actions have literal consequences in a story may undermine a story's integrity. This point is discussed in the following section, *Beyond Branching*.

[3] Dove, p. 281

A simple graph-structure hypermedia experience.

In this model, content is embedded in containers (document, cards, screens) that are explicitly linked to other containers.



A large part of the problem with successfully integrating interactivity and narrative stems from the models for constructing such experiences implicit in current "hypermedia" authoring tools like Macromedia Director and hypertext systems like HTML on the Web. In these systems, the author creates links between pieces of content forming a kind of flow chart or graph structure into which the viewer is placed to navigate. In this sense, the author works directly at the level of the viewer interaction. Thus, Dove's notion of feeling constrained to a predetermined "matrix" of interaction is inherent to the model of hypermedia.

Another inherent limitation of hypermedia is the fact that the resulting structure is static and lacks "state," or capacity to store knowledge. The structure itself captures no sense of the history of the experience nor does it have any built-in competency for presenting itself short of the viewers direct use of links or branch points.

Scaling the Story, Scaling the Telling

Hypermedia systems constructed in this way are extremely difficult to "scale" to include large amounts of content. As each new piece is added, the author must consider the potential linkage of that piece of content to every piece of content already in the system. In this way, adding content is an exponentially complex task.

Furthermore, such systems place the author in the position of effectively pre-thinking all possible viewer pathways through the content. Every "hard-coded" link between two pieces of content in effect freezes the function or intention of the linkage into the structure of the navigation. Thus, the organization of

the story is difficult to scale as any change in "retrieval" functionality necessitates large-scale changes to the pre-coded link structure.

Storytelling Systems

Storytelling Systems present an alternative to simple graph-structured hypermedia. The Storytelling System isolates the author from the process of explicitly linking a story's content. Instead, the system uses descriptions of the content and built-in editing competencies to select and sequence materials dynamically. The Storytelling System provides hooks to make this sequencing process responsive to a viewer. Thus the Storytelling System is a kind of "editor in software" or "narrative engine"—a computer program capable of constructing responsive narratives based on the content and description provided by an author.[4]

The Automatist Storytelling System

The word *Automatist* is used to denote the specific model presented in this thesis. As such, the *Automatist Storytelling System* is one possible approach to building a Storytelling System.

Automatism is a word used to describe a branch of the surrealist movement; it represents the process of creating art based on a kind of "automatic" or unconscious free association. The intention is a "truer" experience as meaning emerges from the interactions of individual expressions rather than from a structure imposed from an "exterior consciousness."

The model presented in the Automatist Storytelling System is similarly content-driven and decentralized. Structure and meaning are considered emergent properties of the storytelling process. Rather than there being a central "conducting" process, sequencing decisions result from the interacting effects of individual material presentations.

The Automatist Storytelling System edits by association. Specifically, the approach uses keywords as a means of indirectly defining *potential links* between materials. During the presen-

[4] Davenport describes the design of "adaptive storytelling systems" in:

Davenport, G. "Seeking Dynamic, Adaptive Story Environments" *IEEE Multimedia*, Fall 1994, pp. 9-13

tation process, keywords function in parallel, pushing and pulling the narrative toward and away from specific pieces of content.

Thesis Structure

The Automatist Storytelling System is presented in this thesis in five stages:

Part 1, ***Interactive Narrative,*** presents a theoretical framework for discussing interactive narrative, provides a critique of "branch-structure" narrative, and establishes a set of five *Fundamental Properties of Interactive Narrative.*

Part 2, ***Approach,*** describes the approach of the Automatist Storytelling System by discussing related and influential research. This section also introduces the keyword-based knowledge representation scheme common to both the ConTour and Dexter systems.

Part 3, ***ConTour,*** describes the design and implementation of ConTour, a graphical demonstration of a simple Automatist Storytelling System. The system represents a potential "back-end" or "narrative engine" for an end-user storytelling system. In addition, the application is a kind of "digital editing assistant" capable of producing "steerable" presentations of keyword-annotated materials.

Part 3, ***Dexter,*** describes the design and implementation of Dexter, a generalized system for browsing collections of documents on the World Wide Web. Dexter represents an application of the principles of the Automatist Storytelling System to the problem of supporting "true browsing" on the web.

Part 5, ***Extensions,*** concludes the thesis by discussing possible extensions of the Automatist Storytelling System. This section also presents scenarios for future applications.

1 Interactive Narrative

Beyond Branching

Marc Canter, father of the computer program now called Macromedia Director, recently presented his CD-ROM *Meet the Media Band* at the MIT Media Lab. While presenting one component of the CD-ROM, an interactive music video where, with the help of the viewer, the lead singer explores various dating options, Canter quickly apologized for the piece having "only sixteen endings."

This notion of "multiple pathway" stories where a high variability of plot is the ideal, is extremely prevalent within the multimedia industry, academia, and popular culture at large.

The view seems rooted in popular conceptions of the ideal narrative experience involving being "inside the story." Brenda Laurel introduces her essay *A Taxonomy for Interactive Movies*, by citing one such conception, Star Trek's Holodeck.[1] In the Star Trek universe, the Holodeck is a large empty room into which a computer synthesizes landscapes and characters. Star Trek characters may enter the space, exploring and generally experiencing their synthetic environment exactly *as if it were reality*.

The inherent problem with looking at depicted representations of interactive experiences is that these representations are embedded in already familiar narrative forms. The ideas behind the depictions often exist merely as devices to further the intentions of the more or less traditional storyteller. Furthermore, while the observation of a character placed in

[1] Laurel, Brenda. "A Taxonomy for Interactive Movies" *New Media News*. Vol 3, No. 1, 1989.

some presumably interactive experience might be entertaining, the literal experience might not.

The Holodeck, which Laurel hails as "the best developed, and incidentally the most realistically doable, model for the interactive entertainment environment of the future," is a case in point. The Holodeck is a microcosm of the Star Trek experience itself. It's a place where the familiar Trek characters are placed in fantastic environments to explore and discover. Within this context, characters become involved in a "micro-narratives" which invariably get entangled in the framing narrative of the ship.

To the viewer, the Holodeck is just another setting for the Trek characters. The events that occur in this fantasy space are presented to the viewer in exactly the same way the rest of the Star Trek "reality" is presented, namely with sets, lighting, actors, and a choreographed camera. Rather than being the ultimate model for an interactive experience, the Holodeck is really the ultimate story device. Thus the Holodeck is to "ideal storytelling" what the Transporter is to "ideal travel"—to get from here to there you simply have the computer "put you there."

The Narrative Prime Directive

In the universe of Star Trek, the members of the Starship Enterprise are bound to the principle of the *Prime Directive*. This fictional doctrine dictates that the crew members' actions must not fundamentally interfere with the cultures they encounter.

Star Trek's explicit premise is the exploration of new worlds and people, "to boldly go..." and so on. To that end, the Prime Directive stands not only to protect the worlds that are explored from being altered by the explorers, but presumably to maximize the meaning of that exploration by preventing such an alteration. In other words, when the actions of the observer significantly alters that which he observes, how can the observer know whether his observations are inherent to the subject or the result of his own actions?

The above situation exists for any observer. Cinema verité filmmakers must accept the fact that their own presence and the

presence of the camera inherently alters the situation they attempt to faithfully record. To this end, the filmmaker is obliged to admit this influence and develop techniques to minimize the significance of its effect.

In storytelling we find a similar "split between worlds." The world created by the narrative, the *diegesis*, and the act which produces it, the *narrating*, are necessarily distinct.

Genette defines the idea of *narrative levels* by stating that "any event a narrative recounts is at a diegetic level immediately higher than the level at which the narrating act producing this narrative is placed." [2] Thus the foundation or bottom-most level of any narrative consists of a narrating act. It is at this level that the intended reader, or *narratee*, receives the narrative. Events which occur within the world of the narrative, or diegesis, are termed *diegetic*. Events occurring at the level of the narrating act are considered *extradiegetic*.

Genette goes on to define any shift between two levels of narrative, such as between the diegetic and non- or extradiegetic, as a *metalepsis*.

In its most straightforward form, a metalepsis occurs when the narrator tells the story of a character telling a story. At that point, the character assumes the role of narrator in a new higher level or meta-narrative. Genette designates any other form of metalepsis *transgressive*. In these cases, metalepses occur unexpectedly as an element initially perceived at one level is revealed to be or treated as if it occurs at another. Genette sites Laurence Sterne's *Tristram Shandy* as a prime example of this latter use.

On Sterne's use of metalepsis in the narration of Tristram Shandy, Genette notes that:

> Sterne pushed the thing so far as to entreat the intervention of the reader, whom he beseeched to close the door or help Mr. Shandy get back to his bed, but the principle is the same:
>
> [Any] intrusion by the extradiegetic narrator or [reader] into the diegetic universe ... produces an effect of strangeness that is either comical ... or fantastic. [3]

[2] Genette, Gérard. *Narrative Discourse*. Cornell University Press, 1980. p. 228

[3] Genette, pp. 234-235

A quintessential set of examples of such an effect in film involves background music. If, for instance, as lush romantic music swells, one character says to the other "Sorry? I couldn't hear you over the music," or in the style of Tristram Shandy, a character asks that the sound be turned down (and then it is), the effect is typically comical. Wim Wenders' *Wings of Desire* presents an example tending more toward the fantastic when music in the soundtrack abruptly ceases as a character, an angel, clasps his hands over his ears. [4]

Summing up the subject of transgressive metalepsis, Genette states that:

> All these games, by the intensity of their effects, demonstrate the importance of the boundary they tax their ingenuity to overstep, in defiance of verisimilitude—a boundary that is precisely the narrating (or the performance) itself: a shifting but sacred frontier between two worlds, the world in which one tells, the world of which one tells.
>
> The most troubling thing about metalepsis indeed lies in this unacceptable and insistent hypothesis, that the extradiegetic is perhaps always diegetic, and that the narrator and his narratees—you and I—perhaps belong to some narrative. [5]

Philosophical concerns aside, transgressive metalepsis are at the very least exceptional and disruptive points in a narrative. The repetition of such effects tends to distract the viewer from the depicted story and undermine the integrity of the diegesis.

"Branch Point" interactions, where the viewer acts to change the course of a narrative, are thus disruptive not only when they stop the story to wait for user input, but fundamentally by placing the consequence of an action performed by the viewer within the diegesis.

By presuming to break the diegetic barrier, such approaches invariably undermine a narrative system, calling into question the boundaries between the narrative, story, narrating act, and ultimately the reader. While such an action may well be a potent tool for the author interested in calling these relationships into question, the technique shifts the viewer's attention from the narrative to its disruption. Furthermore, if such "effects" are perceived as novelties, they cannot form the back-

[4] This reference was suggested to me by Sawad Brooks.

[5] Genette, p. 236

bone of an experience; the novelty wears thin quickly, the comic or fantastic quickly become tedious.

Focusing only on this exceptional potential of interactivity needlessly limits its role to one of disrupting narrative. Instead, by accepting its inherently extradiegetic position, interactivity may be seen, like the background music or editing in film, to *serve* the narrative by focusing the viewer's attention on the product rather than the process of the narrating.

Hollow Stories

In Laurel's taxonomy, one of the three "definitional variables" for describing interactive movies is *personness.* Laurel defines a first-person experience as one where the viewer "participates directly in the world of the movie... [where the viewer] is a character in the action". Second-person is where "the [viewer] stands outside the imaginary world but communicates to characters or entities inside it, or vice versa." In third-person experiences "the [viewer] stands outside the action altogether," a situation Laurel notes "is hopefully never achieved in interactive media!" [6]

Such a breakdown, by its very literal notion of being either inside (good) or outside (bad) the "action" or diegesis of the movie, leaves no place for immersive narrative experiences. In Laurel's second-person experience, every instance of a "communication" between viewer and story represents a transgressive metalepsis—a disruption of the narrative. Laurel's notion of a first-person interactive experience denies the presence of a diegetic barrier altogether. In this case, there is no narrative; the viewer is put into an environment and acts—his or her experience is the story. Removing the viewer from the "action" altogether (third-person) is presented as entirely undesirable as the experience reduces to nothing more than an ordinary cinematic experience, presumably something the interactive moviegoer would not tolerate.

The fundamental problem with Laurel's taxonomy is the implication that one cannot provide an "immersive experience" without literally making the viewer a character. Undeniably, having viewers experience a story *as if they were present* in the diegesis

[6] Laurel

is powerful. At the same time, giving viewers a sense of presence need not bind their actions to consequences in the story.

Discussing the *significance* of interactivity, Laurel poses the question, "Are the user's choices simply cosmetic, or do they fundamentally affect the plot?" [7] In a narrative, this notion of significance seems inversely defined, since the ability to alter events in the plot actually works to diffuse the significance of the story. If viewers can change characters' actions with the wave of their hands, why should they care about the story? What indeed then is the story? [8]

Genette provides a more useful description of the conditions for what might be called *Narrative Immersion* near the end of his essay:

> [The] more transparent the receiving instance & the more silent its evocation in the narrative, so undoubtedly the easier, or rather the more irresistible, each reader's identification with or substitution for that implied instance will be. [9]

In other words, the ideal is not for the viewer to map directly into the story, but rather into the *telling* of that story.

Counterexample: Interactive Theater

In John Krizanc's interactive theater piece *Tamara,* actors performed scenes in parallel throughout the many rooms of a real house. While audience members were free to follow characters throughout the house and observe their performances, the audience's presence was never acknowledged by the performers. Thus the "diegetic barrier" remained firmly in place even though the audience and actors shared the same literal space.

In this scheme, the audience was *continuously interactive,* in that they were free to move their bodies normally. The range of their interaction was quite limited—an audience member could only leave or enter a room when characters did so. The significance of the interaction was high as the author designed the room's performances in such a way that each audience member only gets a portion of the total story depending on what rooms they visit. In this way, audience members are encouraged to

[7] Laurel

[8] This situation is discussed by Glorianna Davenport in the article: "Cinema with 'thinkie' appeal" *Financial Times*, London, Feb. 5, 1996.

[9] Genette, p. 260

speak to each other during a dinner held at the "intermission," to try and piece their disparate experiences together.

I never saw Tamara performed. My understanding of its content and operation comes from informal discussions with the author whom I met at a workshop in Banff.[10] In these same discussions, amid an audience of primarily other writers, Krizanc described his recent discussions with a developer about producing Tamara in CD-ROM form and showed the lengthy design document the developer had produced outlining an approach to the project. Amid concerns over bandwidth, the document suggested that the play be recast as a game with the viewer choosing plot lines resulting in one of many possible endings. For Krizanc, the very idea that his main character, the Count, might live instead of die depending on the actions of the "player" was unacceptable; it "completely destroyed the whole point of the story" as he had written it.

In addition to underscoring the strong negative potential of handing control of events in the story over to the viewer, this incident demonstrates how bound the idea seems to be to a notion of interactivity. Even when presented with a narrative with a ready-made model for interactivity, the CD-ROM developer felt obliged to replace it with a branching narrative structure.

Fundamental Properties of Interactive Narrative

In response to Laurel's taxonomy, this section describes a set of five "Fundamental Properties of Interactive Narrative." As aspects of each property are discussed, any sense of an ideal refers only to the ideal capabilities of a general purpose Storytelling System. In other words, authors should clearly be allowed do whatever they feel best serves their story. The aspects discussed represent the kinds of controls and capabilities a Storytelling System should have in order to facilitate a range of potential uses.

[10] This meeting occurred at the *Interactive Screen* conference held in January, 1995, at the Banff Center for the Arts in Alberta, Canada.

narrative **intention**
narrative **immersion**
narrative **structure**
narrative **response**
narrative **guidance**

Five Fundamental Properties of Interactive Narrative

Narrative Intention

Narrative intention refers to the point of the narrative, or in the case of an interactive narrative, the viewer experience. The narrative intention should answer the viewer's question, "Why am I viewing this story?" and the author's question, "Why am I telling it?" Ultimately, the narrative intention depends on the author and the specific story he or she wishes to tell.

Ideally, a Storytelling System communicates a sense of purpose to viewers while providing authors with a sufficient level of control over those aspects they feel are essential to the telling of the piece. Note that an explicit notion of intention need not imply an explicit or "heavy-handed" articulation of that intention in the narrative. Indeed, an author's intention might be that viewers perceive an *absence* of intentionality to their experience.

Narrative intention is the broadest and most fundamental property of an (interactive) narrative as ultimately the decisions made concerning the other properties are made in its context.

Narrative Immersion

Narrative immersion refers to the quality of the viewer's reception of the narrative. Recalling Genette's definition of the ideal narrative reception, narrative immersion is achieved when the "receiving instance" is transparent, and "silent [in] its evocation." In other words, narrative immersion concerns how well the narrating functions as a background element to the narrative, silently conducting its construction.

In the interactive narrative, immersion relates to how well the Storytelling System engages the viewer in the diegesis rather than in the mechanisms of its construction. To this end, a primary concern for a Storytelling System is how to manage interaction with the viewer without distracting them from or otherwise disrupting the narrative.

Narrative Structure

Narrative structure relates to the form, shape and rhythm of the narrative. As such, narrative structure is inherently a function of the temporal nature of the narrating; it relates to the time of the telling.

Discussing the uses of temporality in narrative, Genette notes that "a narrative can do without anachronies, but not without ... effects of rhythm." [11] In other words, while events in a narrative might be presented in a strictly chronological way, a sense of variation in the pace of the telling is practically indispensable.

Interactivity fundamentally frees the narrative from a fixed and pre-determined structure. This *does not* mean, however, that Interactive Narrative must lack a sense of structure. Indeed, the perceived "flatness" of many CD-ROMs and hypertexts, where the viewer keeps exploring until they either exhaust the system or themselves, tends to emphasize the importance of structure. As Toni Dove likes to say about many interactive experiences, "closure is boredom."

Consider the function of chapters in the novel. Short of providing absolute closure, the existence of each chapter at least provides the means for a satisfying "breaking off" of one's reading experience. As one reads, there exists a motivation to "at least finish off" the current chapter. The end of a well-designed chapter leaves the reader either wanting to read more, or in a position to put the book down with the pleasant sensation of *wanting* to return later to pick up from that point.

A crucial function for a Storytelling System is the ability to manage the narrative structure as it is constructed, to provide a sense of shape, pace, and rhythm to the experience. In addition, structure might include providing a sense of closure to the viewer's experience—even though the story might be ongoing.

In order to provide structure without needlessly limiting the potential for responsiveness to the viewer, storytelling systems must clearly go beyond the use of rigid template structures. Instead, the storytelling system should be capable of forming an *emergent structure* over the course of a particular experience.

[11] Genette, p. 88

To facilitate an emergent structure, some persistence of structural knowledge over the course of the experience is required. At the very least, a storytelling system should be aware of what materials have already been shown to the viewer and avoid repeating them. Sadly, this basic functionality is missing in any hypermedia based on static "state-free" graph structures. In addition, a "structure engine" might take into account the order of previously shown materials to determine how to structure the experience in the future.

In addition to providing structure, a storytelling system might have special knowledge about how to articulate that structure to the viewer. In cinema, conventions for visual transitions have emerged to represent short temporal ellipses with dissolves, larger ellipses with fade outs, and larger possibly conceptual structural divisions with titles. A storytelling system might be capable of dynamically presenting such conventional transitions to articulate structure while maintaining immersion.

Narrative Response

Narrative response encompasses the fundamental question for any interactive narrative, "How does the narrative respond to the viewer?" or conversely, "How does the viewer influence the narrative?"

In a storytelling system, the answer to these questions relies on what types of control or handles the system is capable of providing. It is then for the author to decide exactly how viewers' actions are mapped to these controls.

Narrative response might occur at the level of the diegesis, as when the viewer literally chooses what happens in a story. Such a use however, as discussed previously, may fundamentally work against both narrative intention and immersion. As such, the most useful potential for narrative response occurs in extradiegetic roles: namely in the mechanics of material selection and the conditions of material presentation.

In considering the potential for viewer influence on material selection, we come to the very heart of the storytelling system and its model for making editing decisions. Depending on that

model, a storytelling system might provide a range of editing controls or handles. In the Automatist Storytelling System, keywords are one type of handle. The viewer's actions might then map into tending toward or away from materials described by those keywords.

A less direct type of handle a storytelling system might provide is a notion of the pacing or speed of the story presentation. Mapping viewer action to this control might be simply literal – a slider control for a "faster" or "slower" presentation. It might also be derived from other kinds of viewer interaction. For instance, perhaps the pacing could be based on the amount or level of viewer interaction; as the viewer interacts more, the pace of the story increases.

In terms of how viewer interaction might map into material presentation, the range of possibilities are dependent on the conditions and form of the presentation. For a computer-based presentation, the visual appearance of elements on the screen and the quality and use of sound are both rich domains for exhibiting viewer responsiveness. In a theatrical installation space, qualities like the use of physical space and lighting are potent ways to express narrative response. [12]

Responsiveness need not only refer to a direct and immediate response to a viewer action. The operation of a storytelling system might be responsive to viewers by incorporating accumulated knowledge of their history with the story into the construction of the experience. Thus, to be responsive to viewers in this deeper sense, a storytelling system must maintain persistent knowledge gathered through an accumulation of interaction. Continuing the simple example above, a viewer's use of "keyword handles" might accumulate to form a kind of viewer interest profile on which the storytelling system could base its future material selections, *even in the absence of future viewer interaction.*

Narrative Guidance

Narrative guidance is the culmination of each of the preceding properties; it concerns the fundamental challenge of providing

[12] I was first exposed to the idea of building responsive interactive environments while part of a team working on the "Wheel of Life," a project co-directed by Glorianna Davenport and Larry Friedlander. This project is described in:

Davenport, G. and Friedlander L., "Interactive Transformational Environments: Wheel of Life" chapter in *Contextual Media*, MIT Press, 1995.

narrative structure and responsiveness while preserving narrative intention and immersion.

Tinsley Galyean, in his doctoral thesis *Narrative Guidance of Interactivity,* utilizes the metaphor of a river to describe a guided interactive experience. In a river, Galyean notes, there is a "continuous flow," in which the viewer "steers" and generally never encounters "dead-ends." [13]

Narrative guidance represents the current to the river—the pull of the story. While viewers might have a range of options for interacting at any given moment in their experience, narrative guidance dictates that those of particular relevance should be emphasized. The relevance of a particular interaction should be a measure of its ability to serve the narrative—its intention, immersion, structure, and response.

In the interest of preserving immersion, for instance, options or choices might sometimes be made "silently" by the storytelling system. In this way, the system prevents a disruption to the narrative by isolating the viewer from having to make a conscious decision. Ideally a storytelling system should be capable of providing a *sliding scale of interaction,* from giving viewers a high degree of conscious control to letting them simply "sit back" and watch the story unfold, as they would with a non-interactive narrative.

This notion of level or degree of interactivity might also be a function or manifestation of the narrative structure. Just as a river might consist of calm stretches punctuated by white water—a storytelling system might orchestrate an experience to have highly responsive segments where the viewer is free to explore then transition to fast-paced segments where the viewer is more or less pulled along and left to enjoy the ride (ideally the river metaphor drops off here and viewers are left *not* fearing for their lives).

[13] Galyean, Tinsley. *Narrative Guidance of Interactivity*. MIT PhD Thesis, 1995. pp. 58-59

Two Design Heuristics for Storytelling Systems

A common experience when viewing CD-ROMs seems to be an increasing frustration with having to use the story's interface to "get at" the content. Eventually, if you're actually interested in the content, you just want the thing to "play out by itself."

Indeed, the common click-to-go-forward paradigm for interaction coupled with static graph structure navigation schemes, seem to place the viewer in an adversarial position with the story. Instead of *giving* the viewer their experience, such a scheme requires that the viewer constantly push the story forward, a situation roughly analogous to listening to a narcoleptic storyteller.

The ability for automatic- or self-playout serves as a powerful design heuristic for building a Storytelling System. Designing around the potential *absence* of the viewer requires that a system be built with enough base-level competence to present its content. To add interactivity then poses an interesting challenge as the role and value of the interaction must always be gauged against its absence.

A similar heuristic for designing a storytelling system is something we might call the *Finder Challenge*. A former colleague, David Kung, developed a streamlined technique for evaluating new CD-ROMs (or seedy roms as he called them). Once he figured out the general scheme of the interface, he'd quit the program to view the contents of the CD directly, the way you'd view the contents of a hard disk (on a Macintosh, this outermost application is called the *Finder*). From there, he'd simply open up and view whatever materials he thought might be interesting before finally tossing the CD to a shelf to collect dust.

As with self-playout, the designer of a storytelling system might imagine a base-level functionality of giving direct and immediate access to all of the story's content. Any additional functionality or control given to the viewer must then be gauged against direct access. In this way, the piece must prove its value by enabling a method of construction better than simple random access.

2 Approach

Autonomous Agents and Automatist Storytelling

Decentralized Systems

The Automatist Storytelling System is an instance of a *decentralized system.* In *Turtles, Termites, and Traffic Jams,* Mitchell Resnick describes the operation of "massively parallel microworlds." In the microworld, collections of simple rule-driven entities, or turtles, operate and interact in a controlled environment. Through their interactions with the environment and other turtles, patterns of behavior emerge that are not explicitly represented in the individual turtles' rules. For instance, Resnick describes one example that simulates the food-gathering behavior of ants by "programming" each ant with four simple rules—none of which explicitly refers to the presence of other ants. When hundreds of such ants are placed in a virtual environment containing food, the ants eventually appear to "work together" to relocate food to the ant colony's nest.[1]

For several reasons, a decentralized approach is particularly well suited to the task of building responsive storytelling systems.

Incorporating the presence of the viewer into a decentralized system is straightforward. The viewer may exert influence over the *emergent functionality* of the system the same way any other component of the system does, by altering an aspect of the environment or influencing the operation of other components.

[1] Resnick, Mitchell. *Turtles, Termites, and Traffic Jams.* MIT Press. 1994. pp. 59-68

In this way, the viewer is a "full-fledged" member of the system and consistently integrated into the experience. This contrasts with the model of hypermedia, where the consistency of viewer interactivity depends on the author's consistency of establishing links. When systems are designed by more than one author, or when the base of content grows, maintaining a *consistency of experience* becomes an increasing burden. A carefully designed decentralized system inherently limits this problem.

Autonomous Agents

Introducing the topic *Designing Autonomous Agents*, Pattie Maes describes a shift in Artificial Intelligence research from approaches based on "deliberate thinking" and "explicit knowledge" to ones based on "distributedness and decentralization." She notes how these new approaches avoid the "brittleness" and "inflexibility" of the former by using "dynamic interaction with the environment and intrinsic mechanisms to cope with resource limitations and incomplete knowledge." [2]

In this excerpt from that introduction, Maes describes the idea of *emergent functionality* in systems comprising *Autonomous Agents*:

> One key idea in these new architectures is that of "emergent functionality." The functionality of an agent is viewed as an emergent property of the intensive interaction of the system with its dynamic environment. The specification of the behavior of the agent alone does not explain the functionality that is displayed when the agent is operating. Instead the functionality to a large degree is founded on the properties of the environment.
>
> An important implication of this view is that one cannot simply tell these agents how to achieve a goal. Instead one has to find an interaction loop involving the system and the environment which will converge [towards] the desired goal.
>
> An agent is viewed as a collection of modules which each have their own specific competence. These modules operate autonomously and are solely responsible for [the computation] necessary to achieve their specific competence.

[2] Maes, Pattie. "Guest Editorial" in *Designing Autonomous Agents*. North-Holland, 1990, p. 1

> Communication among modules is reduced to a minimum and happens on an information-low level. There is no global internal model, nor is there a global planning activity with one hierarchical goal structure. [The] global behavior of the agent is not necessarily a linear composition of the behaviors of its modules, but instead more complex behavior may emerge by the interaction of the behaviors generated by the individual modules.[3]

In a later article, Maes describes an approach to programming the mechanical behavior of a robot based on autonomous agents. Decisions about what action the robot should take at any given moment are based on an "action selection" algorithm. In this scheme, the "competency modules" are based on specific actions the robot arm can perform. The applicability or usefulness of each action is a function of the current state of the environment. When an action is selected and performed, its invocation alters the environment, thus influencing the selection of future actions. In this way, a sequence of actions—a plan—emerges.[4]

Instead of using strict Boolean logic for the action selection algorithm, Maes' system relies on the idea of a *spreading activation network*. Modules are invoked when their "predecessors" are sufficiently active; once invoked, activation spreads to a module's "successors." In addition, modules might inhibit or repress the activation of conflicting modules. In a spreading activation network, selection decisions are made simply by picking the most active module at a particular point in the plan. In this way, the system remains highly decentralized as the selection criteria is distributed by the effects of spreading activation.

Automatist Storytelling

The operation of an Automatist Storytelling System exhibits many of the properties of an Autonomous Agent-based system. In ConTour and Dexter, materials and keywords act as modules with an "internal representation" consisting of a list of associated modules; materials are associated with a set of keywords and conversely, keywords are associated with materials. Both materials and keywords spread activation, when invoked, to their associated modules. The resulting interaction of the

[3] Ibid.

[4] Maes. "Situated Agents Can Have Goals". *Designing Autonomous Agents*. pp. 49-70

spreading activation forms the basis of how materials are selected and sequenced. Thus, the resulting structure of the story is an "emergent property" of the interaction of individual material presentations.

Although the approach taken in the Automatist Storytelling System closely conforms to the ideas of Autonomous Agents, it is significantly different than previous applications of this methodology to the area of storytelling.

For instance, in Maes' own subsequent work, agents are applied in the following way:

> Many forms of entertainment employ characters that act in some environment. This is the case for video games, simulation rides, movies, animation, animatronics, theater, puppetry, certain toys and even party lines. Each of these entertainment forms could potentially benefit from the casting of autonomous semi-intelligent agents as entertaining characters.[5]

Thus, research originally developed in the context of coordinating the actions of a robot arm in an industrial environment is translated quite literally to the the idea of planning the actions of virtual characters in a fictive environment. Viewers are considered a part of the environment and thus, as in Laurel's ideal, "inside the story."

Describing the operation of one such system designed around the story of *The Three Little Pigs*, Maes and Rhodes state that:

> In our model, a story emerges from the interaction between discrete, autonomous characters, controlled either by humans or artificial systems. Each character has its own beliefs [and] motivations... [Characters] choose among their possible actions those that most fit their beliefs and motivations at the time.[6]

In these approaches, the process of story construction is viewed as one of generating a sequence of events, or a plot, based on the potential actions of characters with "motivations" while maintaining a global notion of "believability." Under such a scheme, the challenge of constructing "good stories" is a process of creatively expressing a well-formed chain of events.

[5] Maes, Pattie. "Artificial Life meets Entertainment: Lifelike Autonomous Agents" in *Communications of the ACM*. Nov. 1995. V38 No. 11.

[6] Maes, Pattie and Rhodes, Bradley. "The Stage as a Character: Automatic Creation of Acts of God for Dramatic Effect" Presented at the *AAAI '95 Spring Symposium on Interactive Story Systems: Plot and Character*

In the Automatist Storytelling System, the fundamental units of structure are not *events* to be expressed but *expressions* themselves in the form of discrete units of content, or materials. Instead of characters interacting in an environment which is literally the "story world," individual expressions interact in an environment which is the process of the storytelling. In other words, in the storytelling system, what is simulated is not the story but the process of its telling. As such, the approach draws from theories of association and memory, such as those put forth by Marvin Minsky in *The Society of Mind.* [7]

Relative Value Systems

My first exposure to the idea of spreading activation networks was seeing a demonstration of the *GeoSpace* system by Media Lab students Suguru Ishizaki and Ishantha Lokuge. [8] GeoSpace, described by the authors as an "Interactive Visualization System for Exploring Complex Information Spaces," presents a map of Boston and its surrounding cities. When the user requests information about a particular city, say Cambridge, the map gradually animates to emphasize the streets and other information related to that area. If the user then requests information about another city, say Somerville, the system gradually shifts emphasis to the new area while leaving some residual activation on the Cambridge area. In this way, the graphical presentation of information reflects the focus or attention of the user.

In GeoSpace, the presentation is *seamless;* instead of abruptly switching contexts, the system provides smooth transitions. In addition, the effects of residual traces of activation convey a powerful notion of "context preservation." After making several queries, the display provides a sense of the history of the experience with particular emphasis on the recent past.

Much of the power of this visualization scheme stems from the notion of a normalized system. When a user requests information about Cambridge, activation is injected into that city and spreads to related graphical elements. The system translates the amount of activation into the amount of visual emphasis given to that element—its size, brightness, and depth. Every activation value is normalized to the total amount of activation pre-

[7] Minsky, Marvin. *The Society of Mind.* Simon and Schuster, New York, 1986.

[8] Ishizaki, Suguru and Lokuge, Ishantha. "GeoSpace: An Interactive Visualization System for Exploring Complex Information Spaces" *ACM SIGCHI '95 Proceedings,* Denver, Colorado

sent in the system. As in a monetary system, as activation is added to one element, others implicitly *devalue*. Thus, by gradually increasing the value of one or more elements, the entire system responds to stay in a kind of visual equilibrium.

Thus, a key property of a closed or normalized system of values is that an individual value is only meaningful with respect to the larger "containing" system. This principle, which might be called a *Relative Value System*, forms the basis of the ConTour interface and is also used for the "Materials Listing" component of the Dexter interface.

Simple Keyword Representations

Both Dexter and ConTour rely on a relatively simple keyword-based approach to representing their story content.

One if by Clip, Two if by Stream

Schemes for describing video content to a computer have been the subject of research since the dawn of random-access (and now fully digital) video. Approaches may be generally placed in one of two camps: stream-based or clip-based.

In a stream-based representation, such as Marc Davis' *Media Streams*[9], the temporal nature of video is explicitly incorporated into the representation. Description, in the form of keywords, or in Davis' case a set of several hundred icons, are applied to some duration of the video "stream" in order to describe it.

If one imagines the sum of available video in a stream-based system to be a single stream, then a particular description may be thought of as simply a set of durations on this larger stream.

In a clip-based system, video is broken into discrete chunks or clips. Description is then applied to the entire clip. An example of a clip-based system is Ryan Evans' *LogBoy*.[10] In this system,

[9] Davis, Marc. "Media Streams: Representing Video for Retrieval and Repurposing" MIT PhD Thesis. 1995.

[10] Evans, Ryan. "LogBoy Meets FilterGirl: A Toolkit for Multivariant Movies" MIT MS Thesis. 1994.

each clip is described by sets of slot-value pairings such as "Location: Trees" or "Character: Darcy."

The benefit of stream-based systems of representation is the fact that descriptions exist as independent overlapping layers. The start and stop points for each descriptor do not have to match any notion of a clip boundary. In a stream-based system, one could precisely represent a scene with two characters where each character enters and exits the frame independently. For this reason, stream-based systems are particularly well suited to the task of "low-level" editing, where the video content is considered relatively "raw" and a precise level of description is required to make in and out point decisions.

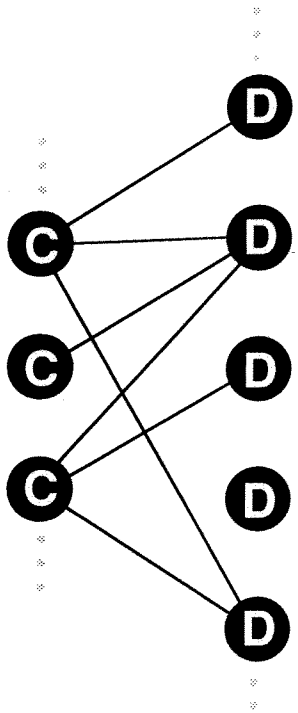
Clip-based annotation, in contrast, provides a further level of control to authors by placing them in the position of structuring the story materials into functional units. Clip-based schemes seem to match the basic filmmaking process of refining "raw footage" by picking out useful or "good shots." Typically, these shots function as a unit; a description scheme that treats them as such is often sufficient. In addition, clip-based annotation may be employed for materials that aren't explicitly temporal such as still images and text documents.

Representation in ConTour and Dexter

In ConTour and Dexter, content is treated as discrete chunks each described by a set of keywords. The relationship between a unit of content and a keyword is neither weighted nor qualified by any notion of "slot" or type. Instead, all weighting occurs only as a function of the presentation; the essential representation remains quite simple.

The *granularity* of each unit of content is a key issue. Each piece must in some sense be self-contained and coherent on its own. On the other hand, due to the simplicity of the representation, a given chunk should only be about one particular set of things. In other words, each piece should form a kind of story *phrase,* complete enough to be coherent, yet not covering too large a range of ideas. Ultimately, the pieces must lend themselves to being dynamically edited together with other related pieces.

In the story applications described in this thesis, video clips tend to be between thirty seconds and two minutes in duration.



In ConTour and Dexter, discrete units of content (materials) are described by units of description (keywords).

By connecting a material to a keyword, the author forms a kind of *potential link* to other materials that are described by or associated with that keyword.

Deferred Sequencing and Extensibility

The essential function of keywords in both ConTour and Dexter is to isolate authors from the process of defining explicit relationships or links between units of content. Instead, the author connects materials only to keywords. By connecting a material to a keyword, the author defines a *potential kind* of connection between the material and others that share that keyword. By connecting each material to a *set* of keywords, the author enables a material to be related to other materials in more than one way.

Lacking explicit links, sequencing decisions are made during the viewing experience based on implicit connections via keywords. Deferring sequencing decisions in this way has two consequences: First, the base of content is truly extensible. Every new material is simply described by keywords, rather than hardwired to every other relevant material in the system. In this way, the potential exponentially-complex task of adding content is managed and made constant. Second, because sequencing decisions aren't pre-coded, viewers may play a more active role in the construction of the experience. Instead of using pre-determined links bound to a specific purpose or organizational scheme, the viewer may influence *how* they want to move from one material to the next.

Hierarchies of Description

One feature of the representation scheme used by ConTour is that keywords themselves may be "described" by other keywords. In *Boston: Renewed Vistas,* this facility is used to group keywords into four "meta-keyword" categories: person, location, time, and theme. Although this facility isn't explicitly available in Dexter, a similar conceptual organization exists for the *Random Walk* keywords.

Materials in these stories are described by a set of particular people, places, times, and themes. As far as the system is con-

cerned, however, the categorizations are irrelevant. Materials may be connected to more than keyword within the same category, useful for instance when a character in the present recalls an event in the past, or when the material addresses several themes simultaneously. Likewise, materials need not be associated with a particular category's keywords at all, useful for "general" clips not necessarily tied to a specific time, place, or character's voice.

By using a keyword hierarchy as opposed to an explicit notion of slot (such as clips having "person slots" and "location slots") the representation is kept as simple as possible. A key point in both ConTour and Dexter is that it's not the representation, but what's *done* with the representation that's interesting and powerful. Furthermore, isolating elements of description in this way is in keeping with a decentralized approach; in this case, slots would needlessly bind components of the description together and disallow their potential usefulness as independent entities.

3 ConTour

ConTour is a graphical demonstration a simple Automatist Storytelling System. The system represents a potential "back-end" or "narrative engine" for an end-user storytelling system.

As an application, ConTour is a generalized system for producing continuous "steerable" presentations of keyword-annotated movies and pictures. In this capacity, ConTour functions as a "digital editing assistant"—interactively suggesting possible sequences of materials. The user steers and shapes the presentation by activating and weighting keywords.

ConTour is the result of several iterations of storytelling systems designed in conjunction with the story *Boston: Renewed Vistas.*[1] This project serves as a model use of ConTour throughout this discussion.

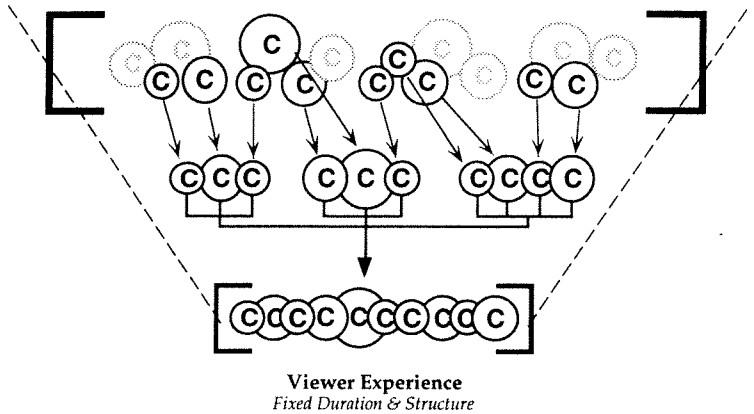
The Evolving Documentary

The "traditional" process of making a documentary film could be roughly described in the following way: The filmmakers collect a large amount of raw material—original film footage, archive photographs, text articles. These raw materials are organized in progressively larger chunks: shots, scenes, and sequences. Finally, sequences are edited together to form the final "cut" of the film. Often this form is in some way constrained in its timing or structure depending on the conditions of its presentation (e.g. television or theatrical release). Regardless, the resulting experience, as presented to the viewer, is rigid and uniform; every viewer sees the same presentation, no matter when or how they see it.

[1] An early version of ConTour, called ConText, is described in:

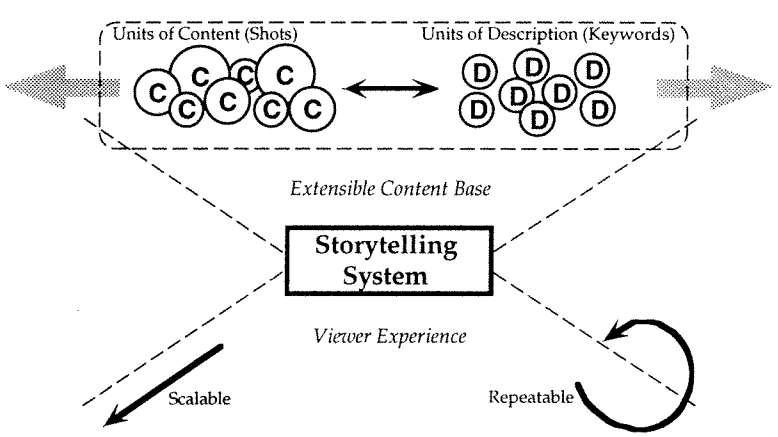
Davenport, G. and Murtaugh, M. "ConText: Towards the Evolving Documentary" *Proceedings of ACM Multimedia '95*, San Francisco, 1995.

Traditional Film/Video Production Model



Described in this way, the filmmaking process may be seen as a kind of funnel, as a large collection of content—frequently an order of magnitude larger in duration than the final piece—is gradually refined and reduced to form the program. As editing decisions are made, such as the decision to place a particular shot or scene at the beginning or end of the piece, the program becomes more and more determined; each placement dictates to some degree the shots and scenes to precede or follow. In this way, as the various pieces fall into place, a particular story, with central characters and themes, begins to form.

The *Storytelling System* Model



An experience based around a storytelling system is more hourglass shaped—open on both the authoring and viewing sides. In this model, the author is isolated from the process of explicitly sequencing their content by the storytelling system; there is no "final cut" of the film. Instead, editing decisions are deferred—made later by the storytelling system in the context of a particular viewing experience.

The viewer's experience is no longer rigid or uniform; the construction of the experience may be sensitive to the conditions of its presentation, including the actions and any available knowledge of the viewer. The experience itself is extensible; viewers are free to stay with the story for as little or as long a time as they wish. The experience is also repeatable; viewers could leave having only seen a portion of the available material and return later to see more.

The system is open-ended on the author's side as well. Instead of "sealing off" the story with the release of a particular program or film, the base of content is free to grow as the story grows. Furthermore, as structural decisions are deferred, the story remains to some degree undetermined and thus free to support presentations with a range of main characters and central themes, as opposed to one particular configuration. For these reasons, we call this form an *Evolving Documentary.*

The Evolving Documentary provides a mechanism for presenting a range of stories that have been traditionally difficult to cover. Specifically, the contemporaneous coverage of stories with long and possibly unknown time spans, as well as stories with a large number of influences and possible perspectives, are particularly challenging for a conventional form like television news. Complaints about television news being too focused on "the moment" and failing to do "adequate followup" seem rooted in the inherent constraints of the form. Examples of stories particularly well suited for an Evolving Documentary investigation are those about wars, urban change, and politics.[2]

The Evolving Documentary form provides an appealing mechanism for developing an "intelligent story archive," allowing isolated materials collected in the present to eventually link to relevant materials added in the future.

Boston: Renewed Vistas

ConTour is the result of several iterations of storytelling systems designed around the story *Boston: Renewed Vistas.* This story, directed by Glorianna Davenport as part of the Workshop in Elastic Movie Time at the MIT Media Lab, concerns the multibillion dollar public works project known as the "Big Dig" cur-

[2] The idea of an Evolving Documentary previously appeared in the context of Gilberte Houbart's *It was a Knowledge War*, an investigation centered on media coverage of the Gulf War.

Houbart, Gilberte. "It Was A Knowledge War." MIT MS Thesis, 1994.

rently taking place in downtown Boston. The project, slated to be finished in the year 2004, represents the largest public works project ever undertaken in the United States and is about 90 percent federally funded. The centerpiece of the project is the removal of the existing Central Artery, a massive elevated highway extension built in the 1950s, and the construction of its underground replacement. The plan calls for the underground construction to take place while the existing roadway remains in operation.

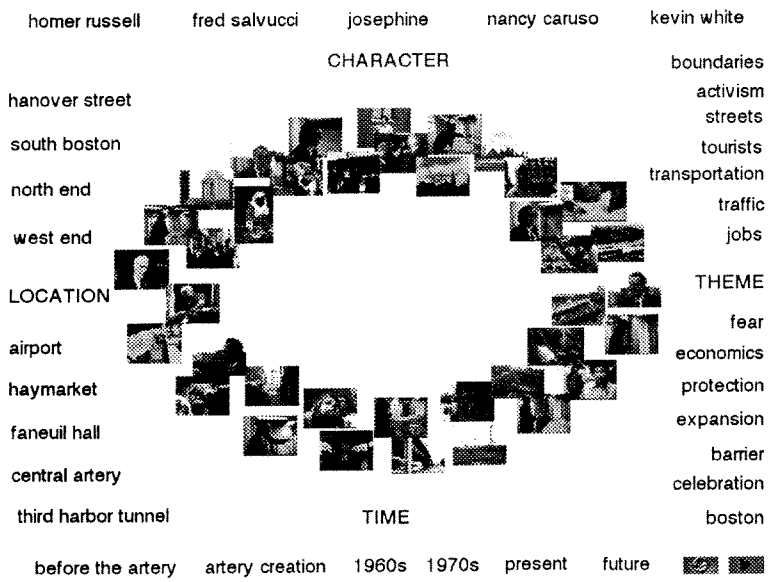
The Big Dig story is well suited to the Evolving Documentary form for a number of reasons. The project is extremely complex and may be seen from a variety of perspectives: the history of the Artery, the politics of how the project came about, the economics of the project's funding, and its impact on adjacent neighborhoods like the North End. The time span of the story is quite extensive and ongoing; the original Artery was built in the 1950s, the formulation of a replacement plan began the 1970s, and its removal will not occur until sometime in the 2000s.

System Overview

The single input to the ConTour program is a text file describing the database of materials. Each line in the file specifies either a text item (keyword), a still picture, or a video clip. Every item also has a text name and screen position. Pictures and video clips are additionally specified by paths to their respective Macintosh picture or QuickTime files. Finally, all items optionally specify a list of descriptors—the names of keywords used to describe that item.

After reading the database, ConTour creates thumbnail images of each of the given picture and video files and displays each item at their given position. Once in ConTour, the user is able to arrange elements onscreen by dragging them while holding the shift key—in its present form, ConTour does not automatically position elements. Once arranged manually on screen, the user may resave the database file with the updated position information.

Two video clips annotated by a set of keywords



Initial screen display of *Boston: Renewed Vistas*

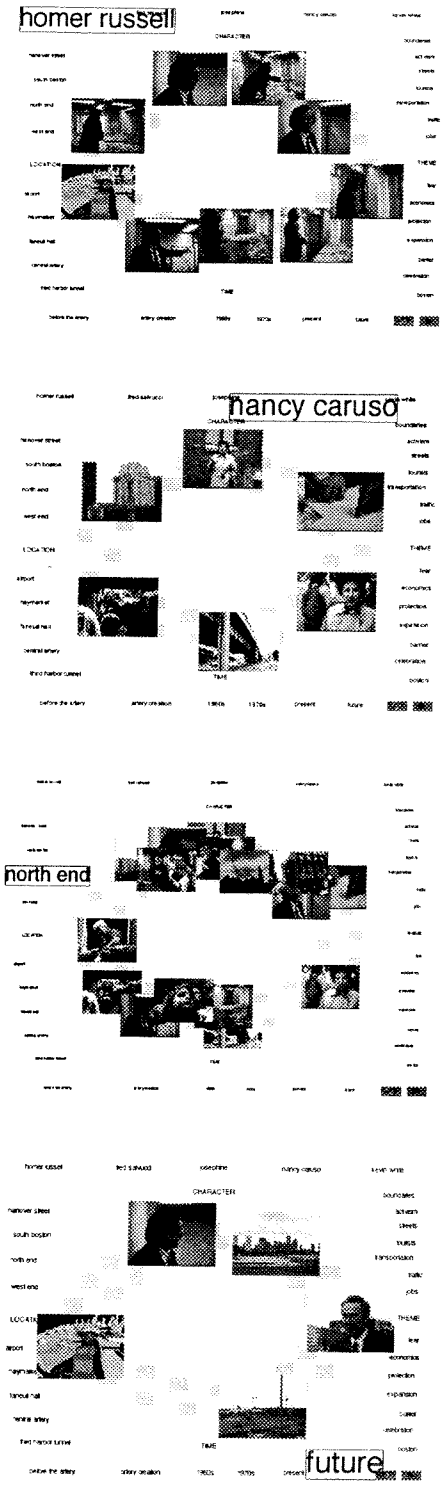
Activation Values and the Graphical Display

Every keyword and material in ConTour has an associated *activation value.* When a keyword is clicked on or a material is presented to the viewer, the element's activation value is raised—the element is "injected" with activation.

Together, the activation values of every keyword and material in ConTour form a closed or *relative value system,* which serves as the basis for both the automatic material selection algorithm and the system's graphical display.

Activation values are used to determine how elements are drawn on the screen; an element's size, depth or z-coordinate, and brightness, are all derived from its activation value. The system uses activation to represent an individual element's relevance to the current "context" of the story playout. Elements with relatively high activation values are made visually prominent by making them appear brighter and closer than elements with lower activation values.

Each element's size, or relative amount of screen space corresponds to its relative amount of activation. Specifically, each element's "screen area" is determined by multiplying the ratio of its activation value to the total amount of activation in the system by the total amount of screen area.



"Database Coverage" views of *Homer Russell, Nancy Caruso, North End,* and *Future*

Each element's depth and brightness are determined in a similar way. Individual activation values are evaluated in the range from the system's minimum to maximum activation value. The minimum value is mapped to the bottom-most position and made least bright while the maximum value is mapped to the top-most position and made most bright. All other values are linearly mapped to in-between depth and brightness values.

In sum, the system uses activation to convey the current *context* or focus of the story presentation; active elements appear prominently in the "foreground" while less active elements fade to the "background."

Two Basic Rules of Operation

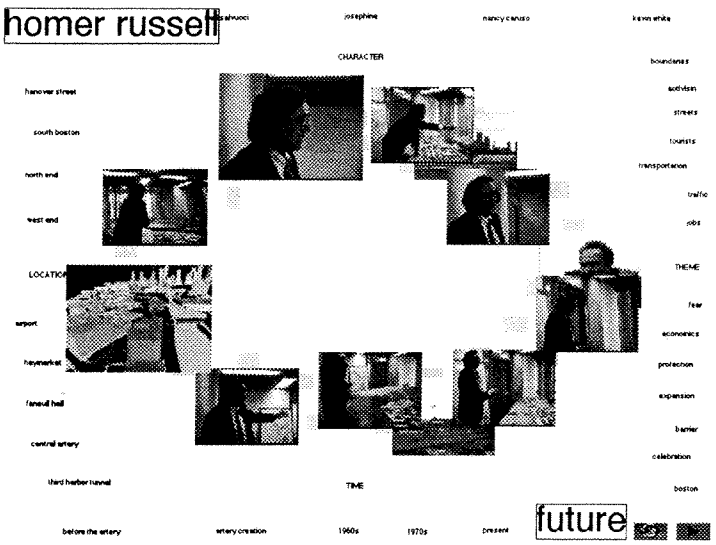
The basic operation of ConTour may be summarized by the following two rules:

> 1. When a keyword is activated, it spreads its activation "downward" to materials described by that keyword.
>
> 2. When a material is presented to the viewer, activation is spread simultaneously "upward" to each of the keywords used to describe the material, in turn invoking the effects of rule 1.

When the user clicks on a keyword, the keyword's activation value is incremented by a relatively large amount. As a result, this activation spreads to each of the materials described by the keyword. In this way, the user is given a graphical sense of that keyword's "coverage" or use in the database.

When more than one keyword is activated, the effect is additive; materials described by both keywords are the largest and float to the top while materials about just one of the two keywords remain slightly smaller and appear farther back. Materials not described by either of the two keywords recede fully into the background. This is an inherent property of the normalized relative value system: as activation is added to certain elements, the relative value of others *implicitly decreases.*

In ConTour, non-text elements below a certain depth threshold are drawn in shades of gray. Visually, the effect is that back-

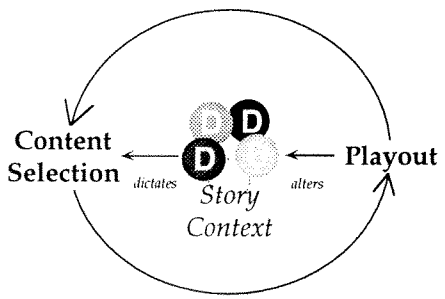


The "additive effect" of clicking on both *Homer Russell* and *Future*

The two largest thumbnails are described by both keywords—the rest by either one or the other.

ground elements go "out of focus," drawing the viewer's attention foreground elements. This threshold, defined as a percentage of the maximum activation value at a given time, may be defined at the top of the ConTour database file—generally the threshold is set at 50%.

The second rule of operation completes the picture. When a material is selected for playout, activation is simultaneously injected into each of the material's keywords. Indirectly, by the operation of the first rule, activation spreads to other materials described by these keywords. In this way, the presentation of a material has the indirect effect of increasing the selection potential of similarly described materials. Once the next material is selected and presented, the process repeats. This property is termed *description feedback.*



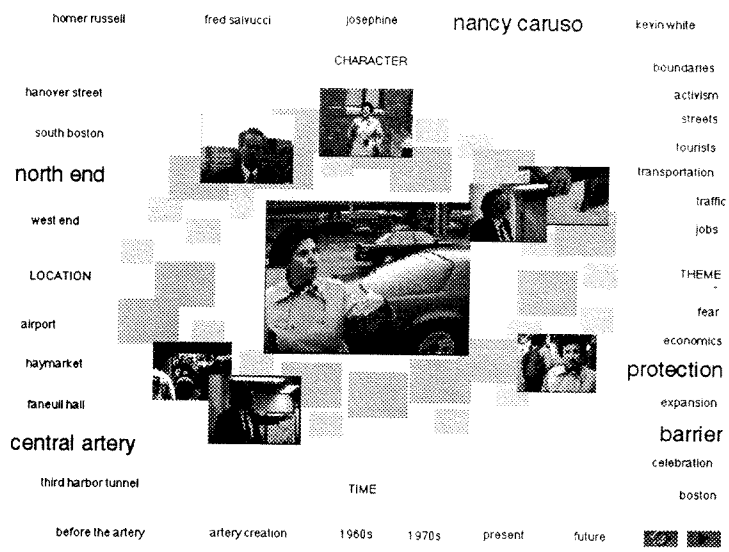
Description Feedback

Automatic Playout

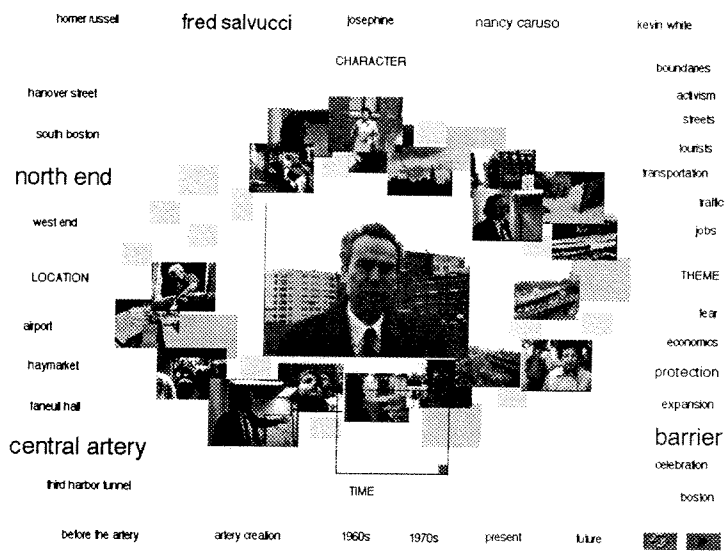
By clicking on the "play" button located in the lower-righthand corner of the screen, the user toggles the automatic playout mode. When activated, the system presents materials continuously, one at a time. The system selects materials for presentation by picking the material with the highest activation value. If there's a "tie" the choice is made at random from among the "high-scorers." As soon as the material is finished, or if the viewer clicks on the material to stop it, the next most active material is selected and presented. The following is an example of an automatically generated sequence of clips:

Playout begins either by a random selection from the database (if no keywords are active and no materials have already been presented), or by the user clicking on a particular thumbnail.

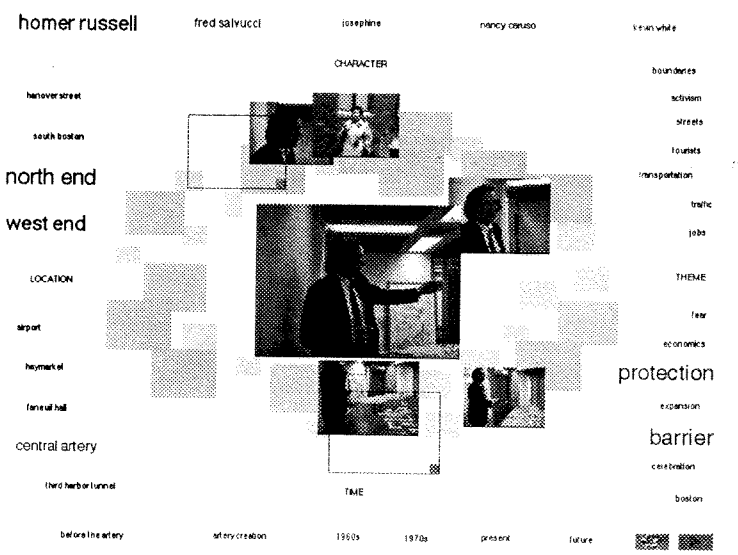
In this clip, Nancy Caruso describes the Central Artery as a "Protective Barrier" for the North End. The clip is described by the keywords: *North End, Central Artery, Nancy Caruso, Protection,* and *Barrier.*

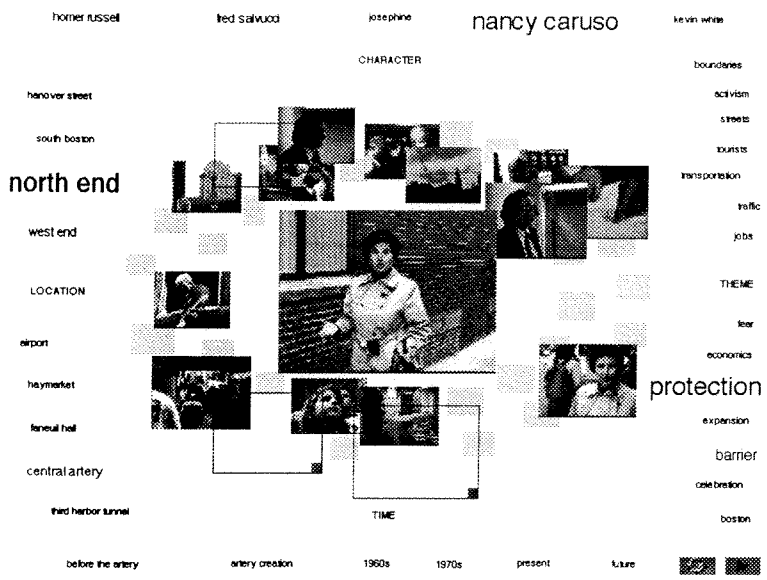


The system next selects a clip where Fred Salvucci describes how the Artery's removal may act to "erase the scar" its construction created. This material is also described by the keywords: *North End, Central Artery,* and *Barrier,* as well as *Fred Salvucci.*



In the next clip, Homer Russell describes the Artery as a "Chinese Wall," cutting the North End off from the rest of the city, and the city from the waterfront. He also refers to Boston's past experience with a neighborhood called the West End.



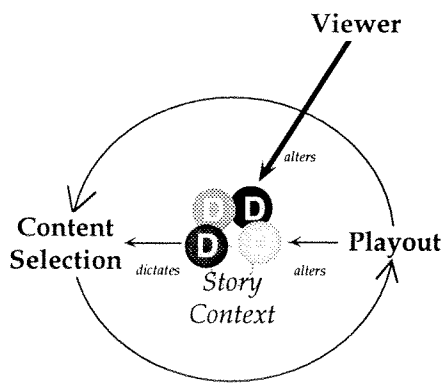


In this clip, Nancy Caruso describes what a "protected community" the North End has traditionally been.

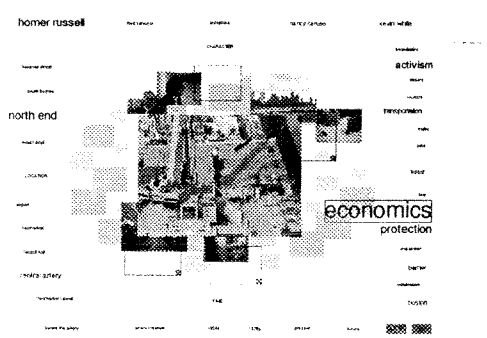
Throughout this sequence, note how the presented clip appears as a prominent thumbnail in the previous step—when the material is very active and about to be selected—and as an empty rectangle in the following step.

The system keeps track of what materials have already been shown to the viewer and makes sure not to select the same material twice. Already viewed materials appear as empty rectangles. In this way, the system keeps "moving forward." By clicking on the lower right hand corner of an empty rectangle, the viewer may manually re-display a material.

ConTour automatically sequences materials based on a kind of "free association" model. As a material is presented, it biases the selection process toward other materials with similar descriptions. When the next material is presented, any new keywords—those not in common to the previous material—get "added to the mix." In this way, ConTour moves slowly through the database of materials in as "connected" and coherent a way as possible.



When the system is in automatic playout mode, the viewer may continue to influence the presentation by activating keywords. For instance, in *Boston: Renewed Vistas*, the viewer might click on the theme "economics" to pull the story in that direction. Due to the systems decentralized selection scheme, the effect of the viewer activating a keyword simply adds to or complements the description feedback process. In this way, viewers may *steer the presentation* toward particular topics of interest.

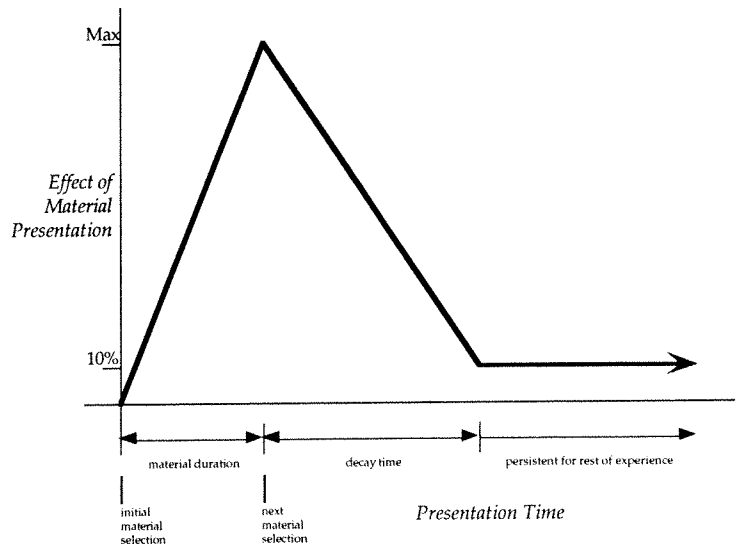


The viewer clicks on *economics* to steer the presentation.

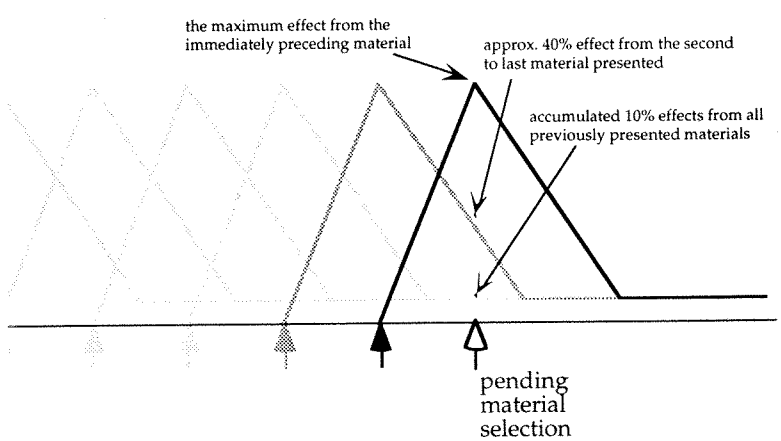
Temporality of the Material Presentation Effect

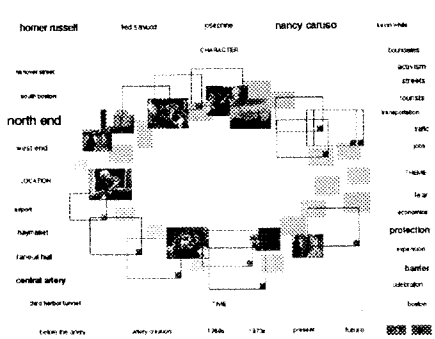
An important subtlety to the effect of material presentation is the way it takes place over time. All materials have a presentation duration. For video clips, it is the inherent duration of the video; for a still picture, the duration is a constant set to something like 5 or 10 seconds.

When the presentation of a material begins, the activation spreading effect occurs gradually over its duration. In this way, the effect of a material presentation is maximal just as the selection of the next material is made. Once a material is finished, 90% of the activation effect gradually dissipates over 1.5 times its original presentation duration. In this way, the majority of a particular clip's effect is restricted to just one or two subsequent editing decisions.



The remaining 10% activation effect from every presentation persists and accumulates over the course of the entire experience (or until the system is reset). This "description sediment" slightly biases the presentation toward keywords the viewer has had some prior exposure to.





The "memory trace" effect results from the persistent effects of all previously presented materials.

When the presentation is stopped, the system will gradually "settle" to a stable state that, due to the persistent effects, exhibits a kind of memory trace showing the degree to which various keywords have been activated.

In sum, the effect of a material's presentation may be seen as having two components, each occurring at a different structural level of the experience. The initial maximal effect is highly localized to a specific material presentation and acts to maxi-

mize the descriptive coherency between individual "shots." We might therefore call this effect a "scene-level" competency. On the other hand, the 10% effects are present for the entire experience—they constitute a "program-level" effect. In terms of its function, the 10% effects accumulate and represent the slowly expanding "scope" of the program—one that represents the viewers' preferences if they've chosen to steer the story toward topics of interest.

Depth vs. Breadth with Spread-Weights

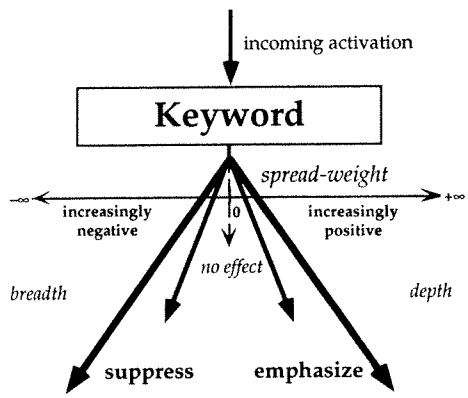
In its basic mode of operation, the description feedback loop is strictly positive; when a keyword is activated, it tends to emphasize its related materials. The result is a *depth-first* exploration of the database. When the user sees a material about a particular theme or person, they tend to see more about that theme or person.

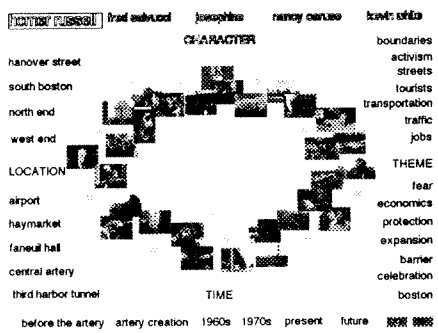
We alter this situation if we allow keywords to exhibit a kind of gain control or *spread-weight.* For instance, we might negatively bias a keyword so that when it is activated, it tends to inhibit or suppress its associated materials rather than emphasizing them.

Graphically, keyword spread-weights are indicated as varying degrees of red (positive), blue (negative), or gray (for zero). When a category keyword's spread-weight is set—such as "character" or "location"—the spread-weight for the entire class of keywords is identically set.[3]

By setting the spread-weight of a class of keywords to a negative value, the system tends to present a *breadth* of content relative to that class. For instance, in *Boston: Renewed Vistas*, if character keywords are made negative, then the presentation of a clip featuring Nancy Caruso temporarily suppresses other content associated with this character, increasing the likelihood of a different character in the next clip.

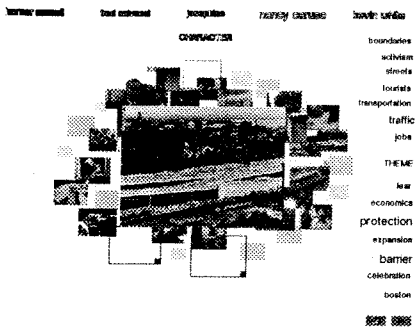
The presentation of materials with zero-weighted keywords has no bearing one way or the other on the selection of other materials. In this way, the user can choose at times to disable certain keywords or classes of keywords. In *Boston: Renewed Vistas*, for





Clicking on *Homer Russell* when characters are negatively weighted causes related materials to be *suppressed.*

[3] In ConTour, clicking on a keyword while holding down the option or command key cycles through one of five spread-weights: very positive (the default), to slightly positive, zero, slightly negative, and very negative.



Here, the system is set to present a *depth of theme* over a *breadth of character.* Location and time keywords have been rendered *inactive* by zero weighting.

instance, we might choose to zero-weight the location class to make locations irrelevant to the selection process.

The full potential of using spread-weights is realized when two or more classes of keywords are set to different values. For instance, by making character keywords negative and theme keywords positive—the resulting playout will tend to present a range of characters' viewpoints focused on particular themes. Operating this way, the presentation could be said to "develop themes." In contrast, the spread-weights could be reversed to "develop characters" instead by presenting the range of themes associated with a particular person.

The addition of keyword spread-weights demonstrates the potential of ConTour's decentralized architecture. Though a relatively simple addition, spread-weights complement the existing selection mechanism to add a powerful new tool to our "story engine"— a generalized means for controlling the *shape* of a story presentation.

4 Dexter

Dexter is an application of the principles of an Automatist Storytelling System to the problem of browsing a collection of interrelated documents on the World Wide Web.

Dexter represents an alternative model to HTML for creating a browseable collection of keyword-annotated documents. Dexter presents viewers with a dynamic graphical interface that supports browsing based on association.

Dexter was developed in conjunction with the story *Jerome B. Wiesner: A Random Walk through the Twentieth Century.* This project serves as a model use of Dexter throughout this discussion.

Motivations

Navigating the Web: Lost at Sea

A common frustration felt as a Web viewer is a feeling of being lost in an amorphous sea of documents. Arriving at a site's home page, one might have some sense of overview; but as one descends link-by-link into the site, it's all too easy to lose a sense of the surrounding context. The viewer wonders: Where am I? How did I get here? What's around me? How does where I am now relate to where I was a few minutes ago?

Web designers have developed techniques to address these problems. The easiest solution is to subdivide the problem. Just as one arranges a messy desktop by creating piles of related items, the web designer collects documents into sections or areas. Each section might be associated with a color or use other graphic elements to convey a sense of "place" to the viewer. In addition, documents can be framed by an overview indicating the current section and providing one click access to the top level of any other. Thus, the "Navigation Bar" is born—a simple map for the weary web traveler.

As the number of documents grows, however, the same problems invariably reappear. At this point, the designer must either create more sections at the top-level or start subdividing existing sections. In either case, the resulting problem is clear. As the complexity of the organizational structure increases, so does the difficulty of both navigating and authoring it. As a result, these navigation schemes tend to work best for shallow hierarchies of a relatively small set of documents.

A fundamental problem with this model of organization is the assumption that a document exists only in a single location. By assuming each section represents a mutually exclusive group of documents, every new section and sub-section comes to further isolate documents within that structure. In this way, the addition of structure serves *to limit* the way documents might be connected, rather than *to increase* that potential.

As in a physical paper filing system, only one organizational scheme or "route-of-access" is allowed. Thus the physical constraints of the metaphorical folder belie the inherent possibilities of hypermedia; whereas the purpose of a physical folder is to encapsulate its contents to *prevent* intermixing, in hypermedia, the possibilities of intermixing are what it's all about.

Browsing, the way it was meant to be

> Generally, the organization of functionality should reflect the most efficient sequence of steps to accomplish the most likely or most frequent user goal.
>
> – Deborah Mayhew, *Principles and Guidelines in Software User Interface Design*[1]

[1] Quote as it appears in the section "Design Supporting the User's Task" in:

Sano, Darrell. *Designing Large-Scale Web Sites.* John Wiley & Sons, New York, 1996. p.113

The word browsing denotes a casual activity. Browsing is all about exploring possibilities, exposing oneself to a range of potential options, and leisurely exploring those of interest. On the web, browsing has been used to describe one's aimless exploration of web pages—as link gives way to link, human browsers on the web don't need to have a particular goal in mind, just a desire to explore and the patience to match their connection speed.

When designing a web site, supporting a viewer's "browsing" becomes a formidable challenge. The design heuristic given above seems in direct opposition to the intention. How does one design for the "most frequent goal" when the viewer might not have one? More importantly, why should an author cater to the broadest possible needs, when ideally the design should respond to the particular viewer at hand; isn't that after all the promise of interactivity?

A web designer at Hearst New Media recently echoed this frustration to me when he recalled how a site they'd designed based on Good Housekeeping magazine was just fine for the person who really "wanted-that-recipe." Speaking each word, he struck his finger to a table to underscore the precision and determination of the imagined cyber-homemaker. Clearly people often only have a general idea of what they want. In this situation, being presented with possibilities to explore at one's leisure is preferable to having to drill down specific levels to uncover something of interest.

As a storyteller, the problem is particularly important. While the viewer might come to a story with a particular interest or disposition, they're most likely several degrees short of specifying a particular material in a collection of dozens, let alone hundreds or thousands. Requiring viewers to make specific choices before they know very much about a story places a wall between them and the story. An effective storytelling system needs to provide accessible hooks into a story for a range of potential viewers.

Jerome Wiesner: A Random Walk through the Twentieth Century

Jerome Wiesner devoted his life to the areas of science, government, education, and humanism. He directed the Radiation Lab at MIT in the 1950s, served as Science Advisor to President Kennedy in the 1960s, was President of MIT in the 1970s, and co-founded the MIT Media Lab in the 1980s. Toward the end of his life, he became an increasingly prominent advocate for nuclear disarmament. After his death in 1994, Glorianna Davenport along with Wiesner's former assistant, Cheryl Morse, began developing a commemorative portrait of Wiesner's life. In October 1995, in conjunction with the 10th anniversary of the MIT Media Lab, the first version of a *Jerome B. Wiesner: A Random Walk through the Twentieth Century* was released as a hybrid Web site and CD-ROM.[2]

The material of the story consists largely of original video interviews with Wiesner's colleagues, family, and friends conducted over the summer of 1995. In early fall, graduate student Freedom Baird edited this video content along with a large collection of still photographs, into over fifty individual video clips; each clip is typically 30 seconds to 2 minutes in duration. Supplementing this video content are several dozen text materials, including Wiesner's memoirs, personal correspondence, essays by contemporaries, and declassified government documents.

In the spring of 1995, I rewrote the underlying system used in *Random Walk*, Dexter, in the web programming language Java, incorporating a number of modifications from the original. The final version is documented here.

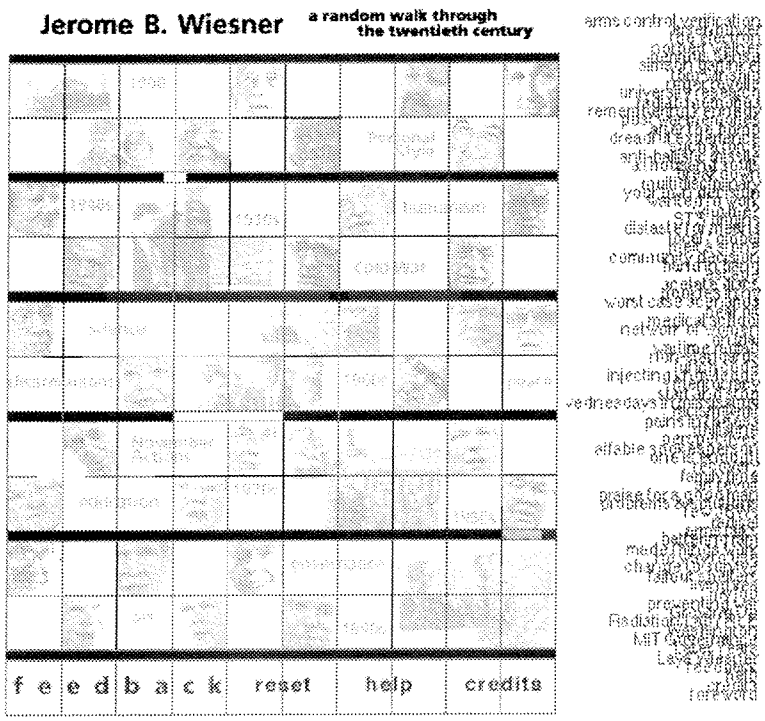
[2] http://ic.www.media.mit.edu/JBW/

Other members of the original team who worked on this project are: Freedom Baird, Ledia Carroll, Peter Cho, Richard Lachman, Geoffrey Litwack, Laughton Stanley, and Phillip Tiongson.

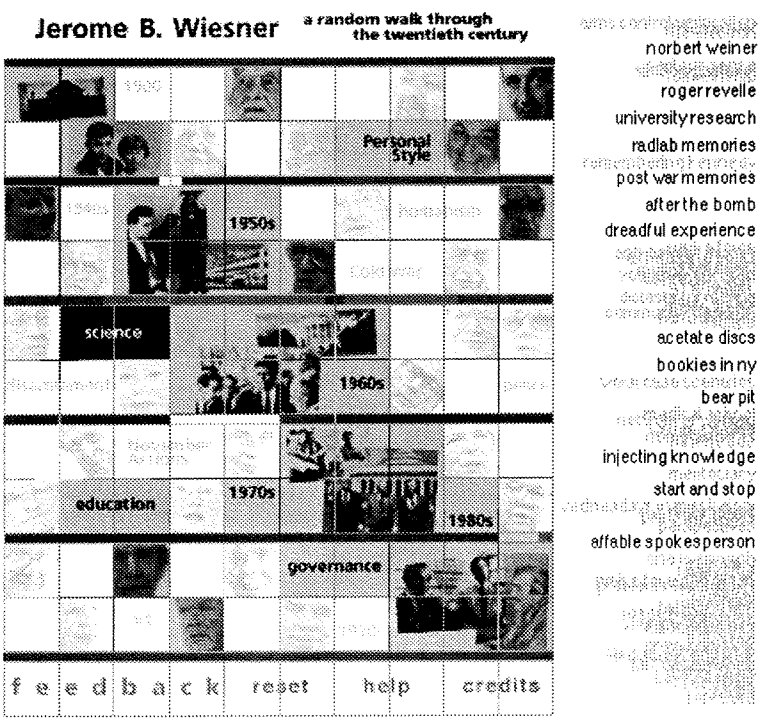
The project was partially funded by the John D. and Catherine T. MacArthur Foundation, Carnegie Corporation, the New Land Foundation, Rockefeller Financial Services, and Schlumberger Limited, Inc.

A Sample Random Walk

This section provides a step-by-step description of one possible viewing of the *Random Walk* site.

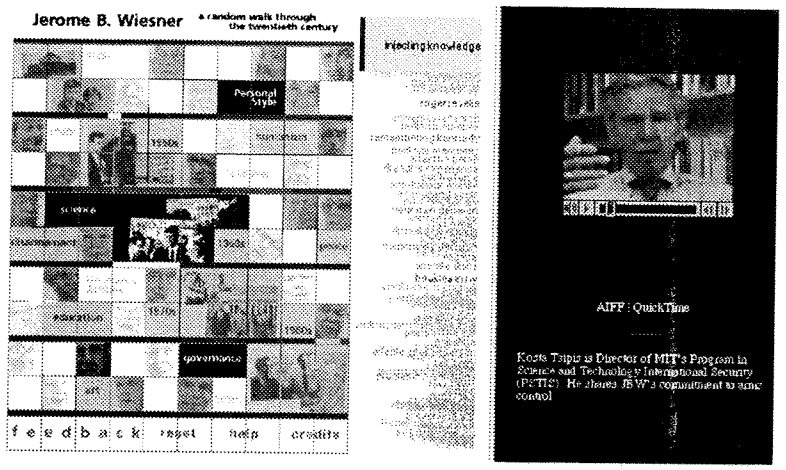


Initially, the viewer is presented with the interface—a white grid of tiles—the *Concept Map*. To its right is a dense vertical listing of text captions—the *Materials Listing*. On the grid are the faint images of people's faces, collages representing five phases of Wiesner's life, and words signifying recurrent themes likes education and science. The adjacent text listing is initially too dense and dim for any individual element to be legible.

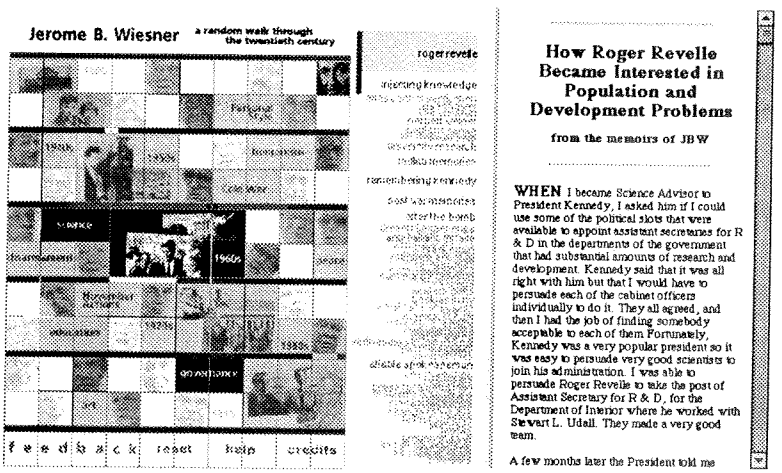


The viewer begins by clicking on *science*; its tiles immediately become active—boldly emphasized in black. In the Materials Listing, the text captions of materials related to science respond, becoming brighter as adjacent elements recede. In addition, other elements within the Concept Map are now emphasized in red—indicating other concepts related (by content) to science. The red squares provide a kind of lookahead, indicating to the viewer those concepts that can narrow their focus from their starting point, *science*.

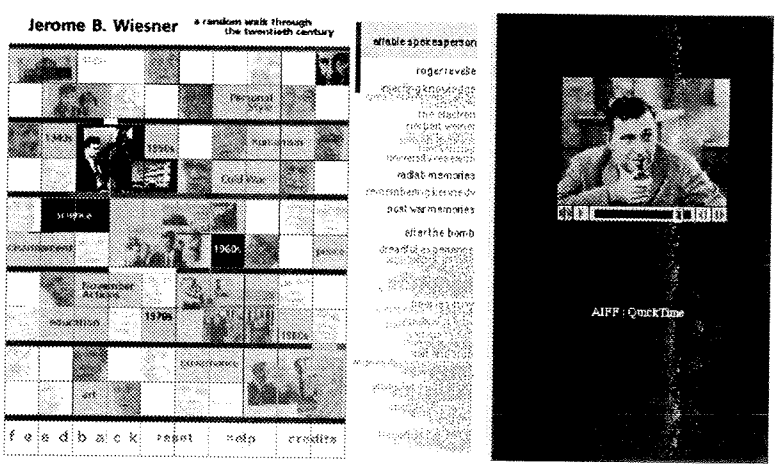
The viewer next clicks on the emphasized tiles marked *personal style*. In response, a material named *Injecting Knowledge* within the Materials Listing becomes highlighted and moves to the top of the listing. In addition to *science* and *personal style*, other keywords related to the selected material become active on the Concept Map. To the right of the interface, the film clip corresponding to *Injecting Knowledge* is presented to the viewer.

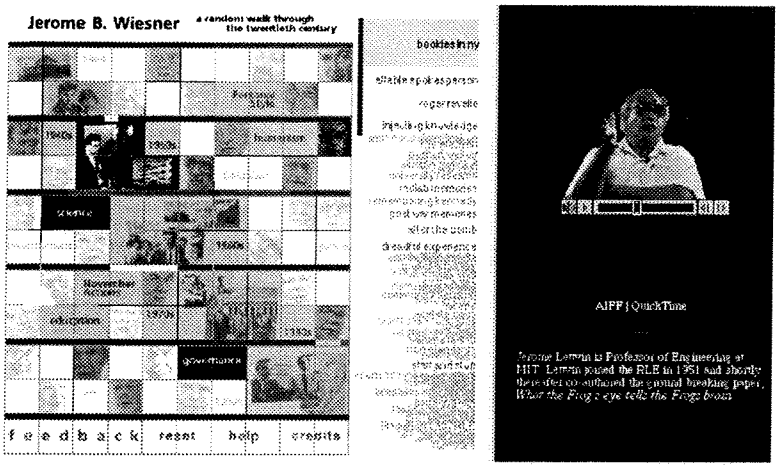


Next, the viewer decides to see more about the Washington time period. Clicking on its tiles—the central collage of images—causes a material from that time period, with as many keywords in common with *Injecting Knowledge* as possible, to be selected—in this case a text document with the caption *Roger Revelle*. The Concept Map again changes to reflect the subject matter of the new material.



Next, the viewer decides to click on the time period preceding the Washington Years. By clicking on the Radiation Lab collage, the material *Affable Spokesperson* is selected—*science, 1960s,* and *Jerome Wiesner* remain active.





Finally, the viewer is interested in an emphasized caption in the Materials Listing titled *Bookies in NY*. Clicking on the caption activates this material directly.

The Materials Listing shows an accumulated history of the viewer's experience thus far.

System Overview

To function as a general-purpose system, Dexter relies on several independent pieces each of which are read into the system when the viewer initially accesses the web site.

The figure on the following page shows an overview of each of the pieces the Dexter system uses to produce the viewer experience. Arrows on the diagram indicate dependencies in the creation process.

Keywords

The first step in the authoring process is determining the set of keywords. By doing this, the author defines all of the possible ways materials in the story may be linked together.

Database (Dexter) File

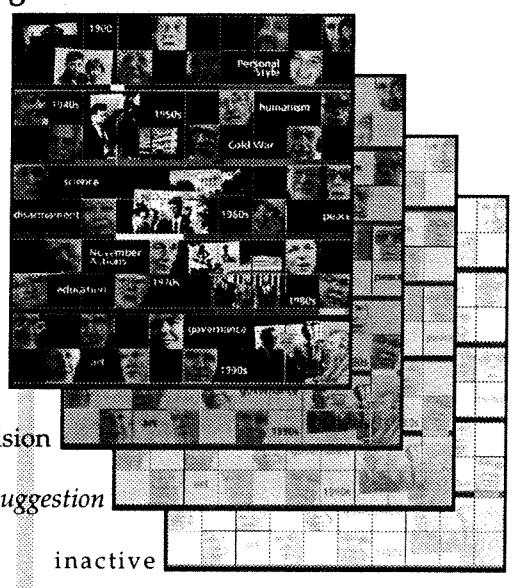
The Database file lists for *each* document or material in the story: (1) a partial URL (relative to the Java applet page) for the "web page" that contains the content, (2) the text caption that appears in the Materials Listing, and (3) a list of keywords that describe the material.

Dexter System Overview

Keywords

People: Amar Bose, Ruth Batson, Laya Wiesner ...
Times: Growing Up, RLE, Washington, MIT, Later Years
Decades: 1950s, 1960s, 1970s, 1980s, 1990s ...
Themes: Personal Style, Education, Disarmament, ...

Interface Images



active

decision

suggestion

inactive

Database (Dexter) File

```
// JBW Dexter Database
// Java Dexter Format

movie {
url="Movies/ABSOULS.html"
name="a thousand souls"
dlist="Personal Style|Amar Bose" }

movie {
url="Movies/ZZFORCE.html"
name="such a force"
dlist="Personal Style|Dottie Adler" }

movie {
url="Movies/ZZANTI.html"
name="anti-ballistic missile"
dlist="Washington|1950s|Disarmament" }
```

Map File

```
// JBW Map File
// Created using WebMap

rect "Later Years" 224,312 320,350
rect "1990s" 192,312 224,350
rect "Nicholas Negroponte" 160,312 192,350
rect "Victor McElheny" 96,312 128,350
rect "Art" 64,312 96,350
rect "Kay Stratton" 32,312 64,350
rect "Later Years" 224,280 320,312
rect "Governance" 160,280 224,312
rect "Emma Rothschild" 128,280 160,312
rect "Kosta Tsipis" 64,280 96,312
rect "Roald Sagdeev" 0,280 32,312
rect "Howard Johnson" 288,242 320,280
...
```

HTML Materials

```
<html>
<html>
<html>
<html>
<html>
<html>
<html>
<a href=ABSOULS.MOV>
<img src=ABSOULS.GIF>
</a>
</html>
```

HTML Materials

Each material referenced in the database file exists as a standard web document, or HTML file.

Interface Images

Dexter reads in a set of four images representing the Concept Map in each of the four keyword states: active, decision, suggestion, and inactive. (these states are described in full below)

Map File

Dexter also reads a standard web "image map" file which, in this case, maps each keyword to one or more rectangular regions on the interface images.

System Operation

The Concept Map

The Concept Map gives a graphical overview of the full set of keywords used to describe the set of documents. As in a geographic map, each keyword occupies a specific region. The visual representation of a keyword may be any combination of image and text—the specific visual design is left to the author. In this way, the appearance of the Concept Map may be tailored to the specific story site. A site might also utilize variations of Concept Maps for different purposes or viewing configurations.

In *Random Walk,* the Concept Map appears as a white grid of tiles, a facsimile of the interior wall of the Wiesner building, home of the MIT Media Lab.[3] Down the main diagonal of the grid, from left to right, appear clusters of images from different periods of Wiesner's life: Growing Up, MIT Radiation Lab, Washington, MIT Presidency, and Later Years. Each time period

[3] The design of the interior wall was done by artist Kenneth Noland as part of a work titled *Here –There,* 1985.

The Wiesner building was designed by I.M. Pei & Associates.

occupies one of the five main rows in the grid; this timeline serves as a rough organizational scheme for the rest of the elements. Adjacent to the time periods are tiles marked with specific decades like the 1940s and 1960s. Beyond these are tiles marked with themes like education, disarmament, and personal style. Filling out the rest of the grid are tiles marked with the faces of the people interviewed.

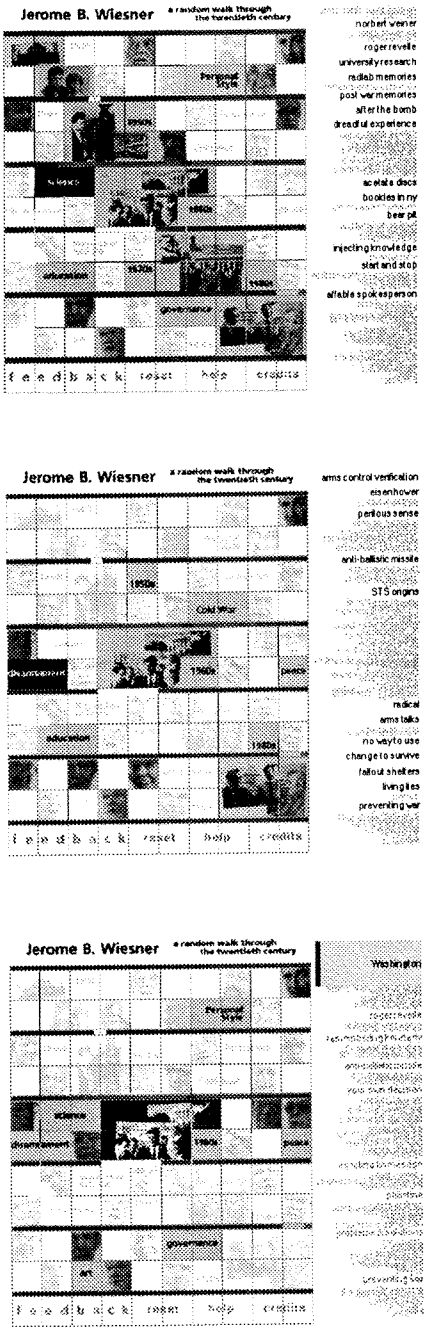
While the position of a keyword on the Concept Map never changes, its visual appearance does change to denote one of four *keyword states*. By controlling the visual appearance of each keyword individually, the system communicates to viewers information about the current state or context of their exploration, and suggests options for navigation.

The four keyword states are: *active, decision, suggestion,* and *inactive*. The state of the Concept Map is determined by the set of active keywords, called the viewing *context*. Initially, all keywords are inactive. By clicking on a keyword, the viewer makes it active thereby changing the context.

As opposed to a web "Navigation Bar" or other button-style controls, keywords on the Concept Map have two additional states other than simply on (active) and off (inactive). These additional states, decision and suggestion, provide a *lookahead*, showing the range of options available to the viewer at each point in their exploration. In this capacity, the lookahead states also provide useful information about how particular keywords are interrelated. [see adjacent figure]

Decision keywords indicate a potential to narrow the current context. Clicking on a decision tile will make that keyword active while preserving all currently active keywords. In this way, Decision keywords allow the viewer to narrow their context by adding another keyword.

Suggestion keywords indicate a partial overlap between the current context and the keyword. Clicking on a suggestion tile will make that keyword active while preserving some *but not all* of the currently active keywords. In this way, suggestion keywords allow the viewer to shift their context to include a keyword of interest while still maintaining some degree of continuity.



"High level views" of, respectively, *Science, Disarmament*, and the *Washington* period

Inactive keywords indicate that no overlap exists between the current state and that keyword. Clicking on an inactive keyword will cause it to become active while making any previously active keywords inactive. Inactive keywords provide the viewer with a means of changing their context completely to begin exploring a new area of interest.

Document Selection

Activation of tiles continues until the set of active tiles uniquely specifies a particular document (or set of identically described documents). For instance, in *Random Walk,* if only one document exists about both Science and the 1940s, clicking on those two tiles automatically selects the document's caption in the Materials Listing and displays it to the viewer to the right of the interface. On the Concept Map, any additional keywords associated with the document also become active. In this way, the viewer's interactions are minimized by removing any unnecessary or redundant clicking.

A second and powerful means of navigation using the Concept Map is the *link-by-concept* mechanism. By clicking on an already active keyword, the viewer requests another document about that keyword with as much in common with the current context as possible (i.e. described by as many of the currently active keywords as possible).

Using this mechanism, the viewer gets to decide how to "move the story forward." Instead of a choosing a particular hard coded link, the viewer chooses to link-by-concept. For instance in *Random Walk,* a viewer might repeatedly click on the keyword *science* to initially explore that theme. As each new document related to science is presented, the system tries to maintain the viewer's context as much as possible. At any time the viewer can start clicking on a different keyword to "switch to a different handle." For instance, exploring materials related to science, the viewer might come upon a document about the Radiation Lab in the 1950s. At this point, the viewer could choose start clicking on the Radiation Lab to develop a picture of that time period and place that is not necessarily related to science.

In this way, viewers are given the means to *structure* their experience around particular keywords while never being bound to a particular organizational scheme. In *Random Walk,* the viewer is free to experience the story as sequences of materials organized by time, theme, or person.

The Materials Listing

The Materials Listing provides an overview of the complete set of documents available to the viewer. Each document appears as a brief text caption arranged in a vertical column. The appearance of the listing responds to changes in the Concept Map. Each caption in the listing is emphasized to indicate its relevance to the current viewing context; the degree of emphasis is based on the number of keywords the caption's document shares with the currently active keywords on the Concept Map.

As a means of navigation, the Materials Listing provides direct one-click access to every document in the system. By clicking on a document's caption, the document is immediately presented to the viewer. When a document is selected in this way, all of the document's keywords are made active in the Concept Map. Thus, the Concept Map also responds to changes in the Materials Listing. By emphasizing document captions based on relevance, the Materials Listing serves as a guide, suggesting documents similar to the viewer's current context.

When a document is activated, either by the viewer's actions in the Concept Map or the Materials Listing, its caption moves to the top of the Listing and is marked with a gray vertical bar. When the next document is viewed, its caption floats to the top and pushes the previous caption one step below it. In this way, the top of the Materials Listing provides a "reverse-history" of the viewer's experience, with recently viewed documents easily accessible on top.

In addition, by marking documents that have been seen and collecting them together, the listing itself becomes a kind of progress bar, showing at a glance what portion of the database has been viewed.

The Materials Listing

All Together Now

Together the Concept Map and Materials Listing provide the viewer with a powerful combination of visualizations and the option to move fluidly between modes of access.

When a particular keyword is activated in the Concept Map, the viewer sees the density of that keyword's use in the Materials Listing as the captions of documents described by that keyword become emphasized. When they select a decision keyword, the viewer sees the context-narrowing effect in the Materials Listings as related elements grow in emphasis while others recede into the background.

By using the Concept Map, viewers take a "top-down" conceptual approach to specifying their interests. Once in a particular context, viewers can use the Materials Listing to select relevant documents directly, choosing captions that catch their interest.

The Materials Listing serves to situate viewers in the context of the larger database. All changes to the listing appear as smooth animations from one state to the next, thus providing a seamless picture of the entire collection of documents. In this way, viewers are always reminded of their place within the "bigger picture."

The Concept Map serves as a memory device, providing a visual means for remembering the "location" of each document as a pattern of Active keywords. In addition, the Concept Map serves as a visual reminder for the way individual pieces relate to others through keywords.

Design Methodology in Random Walk

In Dexter, the appearance of the Concept Map is specified as a series of image layers in each of the four keyword states. In this way, the system effectively isolates the visual interface design from other aspects of the program. Designers are free to produce, using whatever means and tools they wish, the visualiza-

tion of each state rather than having to specify them programmatically. In addition, variations to the interface may be performed and changed without impacting the rest of the system.

The following is a brief discussion of motivations behind design decisions made for the *Random Walk* interface. It also presents a set of general guidelines for designing a graphical interface for use with Dexter.

The visual design of each keyword state needs to make it clearly distinct from each of the other states. Beyond this, certain general guidelines may be made based on each state's purpose and relation to the other states.

Inactive tiles represent those elements not present in the context of the current story—these elements are conceptually "out of focus." In *Random Walk,* the inactive tiles are white with faint gray images and text.

Suggestion tiles are functionally closest to inactive ones. As suggestion tiles play a relatively minor part in viewer navigation, they generally don't need to command much attention. In *Random Walk,* suggestion tiles are light gray—just slightly darker than the inactive state. Visually coupling the inactive and suggestion states also works well since tiles appear most frequently in one of these two states. Thus their visual appearance defines the apparent background of the Concept Map.

A decision keyword indicates the strong potential for that keyword to be made active. The appearance of the decision state might indicate this by a visual similarity to the active state. At the same time, the meaning of the two states is quite distinct; their appearance must therefore be clearly distinguishable.

In *Random Walk,* the active and decision states are both relatively dark in contrast to the lighter inactive and suggestion states. To keep them distinct, however, active tiles are black, the literal inverse of the inactive state, while decision tiles are a deep red. The color red was chosen to suggest the potential for the viewer to act on them; the intention is that by drawing visual attention, we indicate this potential.

In addition to the subtle and relatively minor shifts in coloration and brightness employed in the *Random Walk* design, one could imagine other kinds of changes between states. For instance, the absence or presence of an image in different states could be used to indicate active versus inactive. Shifts in scale or position of image and typographic elements could give an impression of a transformation as keywords change from one state to another.

5 Extensions

Extending the Model

Structure Engine

The keyword spread-weight mechanism described in the ConTour section underscores the potential and flexibility of a decentralized "story engine." However, it also raises the essential question of how such a mechanism would best be used in a storytelling system.

Simply giving control of individual keyword weights to viewers—expecting them to directly manipulate them—makes mechanisms of the story construction highly conspicuous and thus works against providing narrative immersion.

We might instead imagine the author defining the shape of a story by creating a kind of master template specifying the spread-weights of keywords at various times in the viewing experience. In doing this, however, we fundamentally limit narrative response, returning to a pre-determined structure with no capacity for incorporating aspects of a specific experience.

Ultimately, what's needed is a kind of "structure engine"—a system module to manage aspects of the material selection and shape the experience without reducing the system's responsiveness. For instance, in ConTour, a structure engine could be used to manage the keyword spread-weights.

structure engine sketch

Act I—EXPLORE

Attempt to expose the viewer to the full breadth of available content. Accumulate knowledge to build a sense of the PROGRAM SCOPE.

Use "across the board" negative spread-weights for breadth. Use a large number of short sequences for a fast pace.

Act II—DEVELOP

Attempt to expose the viewer to the full depth of content related to the PROGRAM SCOPE.

Prefer progressively longer sequences for decreasing pace.

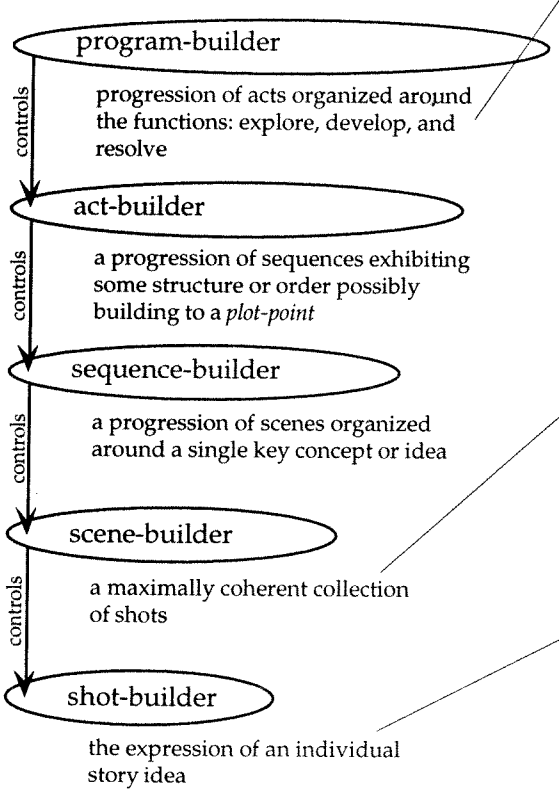
Apply positive spread-weights to members of the PROGRAM SCOPE.

Apply negative spread-weights to all other keywords, or possibly just to members of a CONFLICT SET, and zero-weight (disable) all other keywords.

Act III—RESOLVE

Synthesize closure to the experience by building a fast-paced summary of materials related to the PROGRAM SCOPE that were not presented.

In effect, build a "preview" of possible future experiences based on the viewer's history / interests.



The default operation of ConTour (all spread-weights positive) performs this kind of "scene-level" competency.

The "competency" at the shot-level is performed by the author by breaking their materials into discrete units of expression. Additional information might support variations of the material's presentation.

For instance, a video clip might have several available durations as well as having text equivalents (a full text transcript, a brief caption). In this way, the shot-builder might respond not only to temporal constraints, but to other *conditions of the presentation* (e.g. a low bandwidth internet connection).

In describing the "dual-temporality" of the material presentation effect in ConTour, we see the essence of a simple structure engine already in place. In the short term, ConTour attempts to maintain maximal "descriptive coherency" between individual materials—providing a kind of "scene-level" competency. At the same time, the system accumulates smaller effects from every material presented—maintaining in effect a slowly expanding *scope* of the story.

Extending from this model, we might imagine a structure engine comprising modules, in the style of Maes' Autonomous Agents, organized around a hierarchy of structural competencies: program-builder, act-builder, sequence-builder, scene-builder, and shot-builder.

In the hierarchy, each module exhibits a specific structural competency and only explicitly orchestrates the operation of its immediately subordinate module. For instance, program-builder might produce a program using three instances of act-builder based on the story functions: explore, develop, and resolve.

In addition, each module could be seen as exerting implicit control over every subordinate module by virtue of decreasing orders of magnitude of influence on the selection process. In other words, scene-builder would still operate in a "localized" way to ensure maximal coherency between two materials, but this operation would occur within the larger context and influence of the sequence-builder, in turn under the influence of the act-builder.

In the hierarchy, temporal constraints could "propagate downward." For instance, if the entire program is set to take place over thirty minutes, program-builder might indicate to act-builder that it has ten minutes to construct Act I. In turn, act-builder might use its allotted time to make five two-minute calls to the sequence-builder, and so on.

The final temporal constraint provided to shot-builder might be used to select among temporal variations of a particular shot. These variations might come either from an explicit stream-based representation of possible in and out points, or else might be computationally determined using perhaps computer vision

techniques on the video image or automated pause-detection on the audio track.

A given module might exhibit its specialized behavior not continuously but rather only at specific points in the presentation. For instance, sequence-builder might operate by simply setting up specific story contexts for scene-builder to operate in. In this case, to move to a new scene, sequence-builder would temporarily disable the scene-builder, change the story context to "set the stage" for the next scene, then re-enable scene-builder and allow the presentation to continue.

Other Engines

Other useful engines that could be added to the storytelling system include the following: A "lookahead manager" could use the technique of limited lookahead to "scout out" the ramifications of available options before actually taking them. A "learning module" might incorporate the idea of reinforcement learning and allow pathways through the content to become "burned-in." The authoring model for this kind of system would consist of setting up a loose structure and then *rehearsing* the system until its performances were satisfactory.

Viewer Model

An extremely powerful component not directly addressed in this thesis is an explicit notion of a viewer model. By incorporating a dynamic model of the viewers knowledge over the course of an experience, a storytelling system could utilize concepts like suspense and surprise, in general making decisions based on "what the viewer knows" at a particular point in the narrative—a key property in narrative genres like the mystery as well as in most documentaries.

Certain aspects of "viewer knowledge"—for instance, what materials have been presented—are present *implicitly* in ConTour. Making such knowledge part of an *explicit* and independent viewer module, however, enables that model to be stored after a particular experience then reused in future story presentations. For an Evolving Documentary, such a persis-

tence of viewer knowledge would be essential to make the viewer's experience truly repeatable.

Future Forms

Story Scrapbooks & Spreadsheets

In addition to serving as a kind of "narrative engine" and back-end to a storytelling system, ConTour may be seen as the forerunner to a family of tools that add value to collections of video and other media by allowing viewers to play with variations on their presentation. We imagine two such applications, the Storytelling Scrapbook and the Story Spreadsheet.

In the Storytelling Scrapbook, family members store photographs, text, and video snippets documenting the lives of family members. The Storytelling Scrapbook then supports viewing the "family archive" in a variety of ways; presentations might be organized around a specific event, or trace a recurring event or family member though time.[1] In a network environment, such a tool could facilitate the publishing of forms like the "video postcard"—a self-presenting collection of materials that one sends to friends or distant family members.

As a business desktop application, one could imagine a kind of Story Spreadsheet—a system that lets the user play "what-ifs" with an annotated media archive. For instance, one could imagine conducting market research by collecting a large amount of "testimonial" footage of consumer response to a range of products. These materials might then be "plotted" as a temporal presentation across a variety of demographic axes.

[1] Amy Bruckman describes a range of such scenarios in "The Electronic Scrapbook: Towards an Intelligent Home-Video Editing System" MIT MS Thesis. 1991.

Scaling the Audience

The explosive growth of the internet and the World Wide Web suggests the storytelling system's potential to support not only scalable story content and telling, but also a scalable audience.

A key challenge becomes how to provide experiences that are responsive and meaningful at a personal level while at the same time are perceived as shared across a "society of audience." Currently in the *Random Walk* site, viewers may select a special "feedback" button to email the editors additional stories and related web URLs. In addition to this simple input mechanism, one could imagine a fully integrated "discussion" forum where viewers leave comments and communicate to others. Each communication could be indexed and accessed the same way other content in the site is.

In an imagined Dexter-like system, when viewers submit comments, the messages would be automatically tagged by the current story context. Viewers might also have the option to use the Concept Map to specify a new combination of keywords not already present in the database. One could imagine a second "discussion" materials listing appearing parallel to the "authored" listing. When keywords are activated on the Concept Map, both listings would simultaneously respond providing the viewer with a sense of both the topic's coverage in the database and audience response. Keeping the two sets of materials separate enables viewers to decide how much "outside commentary" they want to incorporate into their experience.

Extending from this example, one could imagine a whole new form of story somewhere between the documentary and the internet newsgroup. In the StoryGroup, the "author" plays the role of moderator and editor, constructing and maintaining the conceptual infrastructure of a story's site while allowing outside reaction and contribution. An author might choose at a certain point to relinquish control completely, allowing the story form to grow based purely on its collective and evolving audience.

Bibliography

References within this thesis are indicated by page numbers in brackets.

Bruckman, A. "The Electronic Scrapbook: Towards an Intelligent Home-Video Editing System" MIT MS Thesis. 1991. [73]

Davenport, G. "Seeking Dynamic, Adaptive Story Environments" *IEEE Multimedia,* Fall 1994, pp. 9-13. [15]

Davenport, G. and Friedlander L., "Interactive Transformational Environments: Wheel of Life" chapter in *Contextual Media,* ed, Barrett and Redmond, MIT Press, 1995. [27]

Davenport, G. and Murtaugh, M., "ConText: Towards the Evolving Documentary" *Proceedings of ACM Multimedia '95,* San Francisco, CA. 1995. [41]

Davis, M. "Media Streams: Representing Video for Retrieval and Repurposing" MIT PhD Thesis. 1995. [36]

Dove, T. "Theater without Actors—Immersion and Response in Installation" *Leonardo,* Vol. 27, No. 4, pp. 281-287, MIT Press. 1994. [12, 13]

Evans, R. "LogBoy Meets FilterGirl: A Toolkit for Multivariant Movies" MIT MS Thesis, 1994. [36]

Field, S. *Screenplay: The Foundations of Screenwriting.* Dell Publishing. 1994.

Galyean, T. "Narrative Guidance of Interactivity" MIT PhD Thesis, 1995. [28]

Genette, G. *Narrative Discourse: An Essay in Method*. Cornell University Press, 1980. [11, 19, 20, 22, 25]

Griffith, V. "Cinema with 'thinkie' appeal" Financial Times, London. February, 5, 1996. [22]

Houbart, G. "It Was A Knowledge War." MIT MS Thesis, 1994. [43]

Laurel, B. "A Taxonomy for Interactive Movies" *New Media News*, Vol. 3 No. 1, Winter 1989. [17, 21, 22]

Lokuge, I. and Ishizaki, S. "GeoSpace: An Interactive Visualization System for Exploring Complex Information Spaces," *ACM SIGCHI '95 Proceedings*, Denver, Colorado. 1995. [35]

Maes, P. ed. *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back.*. North-Holland. 1990. [32, 33]

Maes, P. "Artificial Life meets Entertainment: Lifelike Autonomous Agents" *Communications of the ACM*, Nov '95, V38, No. 11. 1995. [34]

Minsky, M. *The Society of Mind*. Simon & Schuster. 1986. [35]

Resnick, M. *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds*. MIT Press. 1994. [31]

Sano, Darrell. *Designing Large Scale Web Sites*. John Wiley & Sons, New York, 1996. [54]