# Generative Storytelling for Information Visualization

**Pedro Cruz and Penousal Machado**
*University of Coimbra*

In an era characterized by the flow of massive amounts of data, information visualization plays an important role in synthesizing data, making it intelligible, and providing insight from it. Here, we focus on telling a story from data by interpreting events, actions, and actors as a way to amplify the cognition of information. A story uses real-world metaphors and analogies to convey knowledge in a ludic, compact way. So, storytelling, in the context of this article, deals with the core of information visualization by extracting relevant knowledge and enhancing its cognition.

In particular, we address creating stories from *data fabulas* (which we discuss in the next section), using computer graphics as a narrative medium. We've developed a conceptual framework for information visualization: *generative storytelling*. This approach aims to build various stories conveying the same fabula from a given dataset. To illustrate it, we use a visualization of the decline of the major maritime empires.

## Data Fabulas

The four dimensions of reality are constantly being mapped to dimensionally constrained media (for example, 2D maps depicting celestial bodies' movements). Computer graphics can overcome major constraints when mapping space-time by simulating these four dimensions. The intrinsic temporal nature of *n*-dimensional spaces is frequently present in datasets—for example, historical events, financial data, biological behaviors, and model descriptions in physics. These datasets' semantic relevance, together with their temporal nature, seem to beg for a story.

Following Mieke Bal,[1] we define a *fabula* as a set of time-ordered events caused or experienced by actors. In this context, *actors* are agents that perform actions in the fabula's time span. We focus on data fabulas: the set of events, agents, actions, and chronology extracted from a dataset. (In the rest of the article, we use "fabula" to mean "data fabula.") For simplicity, we consider only one extracted fabula per dataset—this implies that the set of events is immutable.

You can present a fabula in different ways. A *story* is the fabula's presentation layer; it has its own structure, built from the fabula's events, actors, and actions. This structure also defines the emphasis given to the events, actors, and actions. So, a story's structure is a *narrative*. The narrative's chronology frequently varies from that of the fabula in its order, pace, or rhythm. The agent that transmutes the fabula into a story is the *narrator* or *storyteller*. This transmutation is called *storytelling*. The narrator communicates the story through some narrative medium (see Figure 1).

## Generative Storytelling and Data

We consider a dataset as a collection of temporal measurements from the real world (data). The changes in these measurements are of utmost importance for storytelling. For every dataset, we can always find a time interval in which no changes occurred for each measurement, owing to

- the system's noncontinuous nature,
- the restrictions in data acquisition processes that limit the frequency of storage to a maximum, or, ultimately,
- the imposition of a threshold below which a measurement's variation is considered irrelevant.

That being said, we can assert that a dataset is inherently discrete. Furthermore, this discrete nature is prominent in most datasets, in which the time stamps of measurements' changes are noticeably spaced in relation to the overall time span.

Usually, the extraction of a fabula from a dataset results from data mining. Clustering and

 Published by the IEEE Computer Society 

aggregating the data play a central role in establishing a chronology, a set of events, and a set of actions, and in identifying the relevant actors. Because the dataset is discrete, we can extract all the changes in its measurements and build a set of time-ordered changes. To construct a fabula, we map these changes' time stamps to events. The fabula's actions represent the values' variations. We associate each event with a corresponding set of actions, modeling how the event affects the actors—an event triggers a set of actions by the actors.

Determining the actors usually doesn't rely much on searching. The dataset's semantics, along with the original ideas regarding the knowledge to extract and portray, play a key role in this process. The data scientist performs these tasks and tailors the clustering and aggregation algorithms' parameters; this makes her or him the fabula's author.

### Generative-Storytelling Characteristics

Because a story is a fabula's presentation layer, it involves

- the representation of the fabula's actors and
- the definition of a temporal structure that, while containing the same events as the fabula, establishes a new pace, rhythm, or chronology, therefore creating a narrative.

These characteristics form the story's identity. Generative storytelling builds the story by implementing the representation of the actors and the new structure of events. The generative nature of the actors' representation emerges from a set of rules that interprets the fabula's actions.

Although the fabula, like the dataset, is discrete, we can portray it in a continuous fashion. We're interested in transforming a discrete fabula to a story with a continuous timeline. This continuity is one of the compelling characteristics of a story. It requires the portrayal of time intervals between events and gives room to the continuous temporal representation of the fabula's actors. Because we're in a continuous environment, the actions' interpretation defines the actors' transitions along time, and, more important, during the absence of events.

The actions' interpretation might also lead to variations of the actors' behaviors derived from the ones the fabula suggested. We can attain this, for instance, by allowing a nonrigorous representation of measures, incorporating behaviors that distort the fabula, introducing peripheral actions, or creating new interactions between the actors. Such *secondary actions* exist only in the story; they're an ornamentation resource that promotes the story's dramatic value without changing its semantic content. So, the story doesn't have to be an accurate representation of the dataset, and the mapping of the fabula's actions to the story isn't linear, fixed, or even deterministic. This is, in our opinion, one of generative storytelling's most distinctive and prominent features, constituting an important narrative resource.

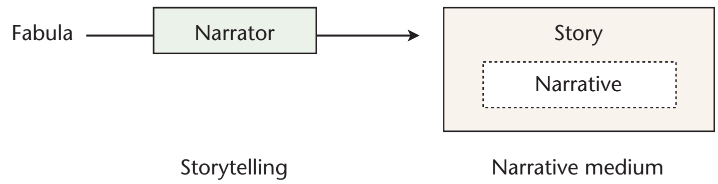


**Figure 1. Storytelling transmutation by a narrator. A story is the presentation layer of a fabula (a set of time-ordered events caused or experienced by actors). The storyteller builds a narrative from the fabula, conveying it as a story through a medium.**

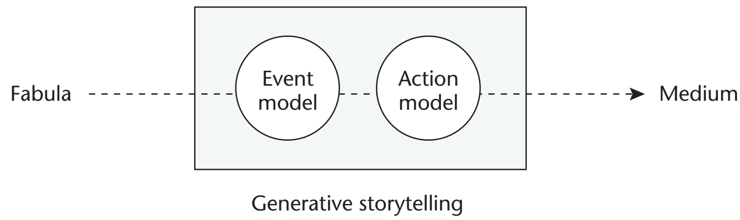


**Figure 2. The generative-storytelling engine. The actors aren't interpreted or manipulated by a model, so they're unchanged by the fabula's transformation.**

### The Conceptual Framework

Our approach relies on an engine that transforms a fabula into a story. This engine consists primarily of two models (see Figure 2).

The *event model* manipulates the time of the events, creating the story's timeline. It processes the fabula events and triggers the respective actions. Although the story is delivered through a continuous medium, the continuity of time doesn't imply linearity. Usually, a story presents a temporal manipulation of its fabula to achieve a dramatic tone. These manipulations not only can change the pace and rhythm but also can alter the chronology itself by introducing narrative devices such as analepses, prolepses, ellipses, in medias res, and reversions.

The *action model* deals with the representation of the fabula's actions, implementing a set of actors' behaviors. To better generalize and define this model's purpose, we consider it as an adaptive system. This means that as the actions are triggered by events through time, the system tends to adapt to those actions instead of reacting directly to them (new actions can be seen as time-based disturbances to the system).

This results in a set of important parameters that we use to adjust the model. One such parameter is the system's adaptation threshold, which determines which actions are relevant. Another is the system's fidelity to the fabula; that is, the adaptive system might promote a loose or tight representation of the fabula. A third is the system's convergence rate to new equilibrium states; that is, the system might react rapidly or slowly to new actions.

### *Room for Expressiveness*

Traditional storytelling media (for example, text, speech, and images) leave room for expressiveness. In our opinion, this expressiveness is the story's foundation, using the manipulations of the fabula as narrative resources. A different expressive intent can thus generate a different story.

Our framework applies this idea to computer graphics, enabling manipulation of the fabula's representation and timeline. The generative-storytelling conceptual framework embraces room for expressiveness with its two models—with the same implementation, you can achieve different expressions by adjusting the model parameters. Furthermore, varying expressions can also result from the same parameters, using a nondeterministic model. This approach maintains the strongest behaviors resulting from a direct interpretation of the actions, with a set of secondary behaviors providing unique expressive depth to each story.

As we previously stated, the resulting story doesn't necessarily precisely represent the dataset. However, the liberty taken in the generative approach results in meaningful variations, adding a dramatic or aesthetic tone and revealing a story's ludic nature. We see this expressive playground promoting cognitive amplification in the context of information visualization. Ultimately, we aim to attain that delicate balance between data portrayal and expressiveness. Assuming that compromises are necessary, we're biased toward data accuracy to avoid corrupting the fabula. The tailoring of parameters should enhance the extraction and portrayal of relevant knowledge without distorting or occluding it.

## Applying the Framework

Here we describe a series of artifacts that illustrate our storytelling framework and discuss their ability to convey the data in a way that's ludic and enticing.[2] We aggregate data, extract a fabula, and build a system that acts as our narrative agent.

### *Visualizing Empires Decline*

We began with an idea for a story: the Western empires' decline in the 19th and 20th centuries. To confer a dramatic tone to the visualization, we chose the top maritime empires. We're acquainted with their glory days and current situations, so we expect vast dimensional changes during the visualization's time span. In contrast, we weren't familiar with the timeline of each empire's growth and decline, which is a vital part of the story. To convey the empires' extension, influence, and relevance, we portrayed the land area each empire possessed over time. We find this measure a well-founded representation for the perception of an empire's influence.

To narrow the contextual spectrum and avoid problems such as visual cluttering and imperceptible proportions in a linear scale, we focused on the four major maritime empires of the 19th and 20th centuries: the British, Spanish, second French colonial, and Portuguese empires.

### *Fabula and Representation*

The dataset consists of roughly 110 events of land gain or loss. Land gain corresponds to conquests, colonization, or acquisitions; land loss results from the independence of former colonies or dominions. So, our measures are the land area and geographical position of each empire and of the new nations. (Each empire is mapped on the simulation canvas to its capital's geographical position.)

Extracting a fabula from the dataset involved a sorting algorithm to establish the events' chronology. The determination of actors was straightforward—each empire or former colony or dominion is an actor. Each gain of land or independence is an event with its associated year, the affected actors, and the actions involved. The fabula contains three types of actions:

- an area increase of the empire, which is triggered by a growth event;
- a division action when a former colony detaches itself from the mother empire, which is triggered by independence events and which decreases the empire's area; and
- a traveling action in which the former colony or dominion moves to its average geographical position, which is also triggered by independence events.

These actions' interpretation is closely related to the actors' representation. We represent each actor as a circle with an area proportional to its land area (see Figure 3). Empires' circles are filled with the color associated with the empire. Former colonies are unfilled circles with rims in the corresponding empire's color. We color the rim instead of the entire circle to avoid visual clutter (after all, we have 95 actors) and emphasize the empires.

We map an area increase directly to the area of the corresponding empire's circle. Division actions involve the emergence of a new nation from the corresponding mother empire, introducing a new actor in the story's timeline. With the appearance of these actors, which will take a rather passive role, we can portray how much of the world was once part of a maritime empire. Each division action involves a secondary action: a circle appears in the empire's circle and starts growing seven years before an independence event, until it reaches the area of the nation-to-be in the year of independence (see Figure 4).

This secondary action exists only for dramatic purposes and to address the need for expressiveness in generative storytelling. Additionally, although territory losses have high visibility (due to the emergence of new unfilled circles), empires' growth is subtle by comparison (the circles simply expand). This lets us emphasize the empires' decline.

### Data Accuracy and Expressiveness

The circles look and behave like soft bodies, colliding with each other and presenting moderate shape variations. This representation involves a physics system that implements the action model, which we discuss later. We chose this behavior to convey the volatility and instability of the empires' evolution through "jigglier" deformations (see Figure 5).

The collisions represent the empires' aggressive quest for hegemony. So, former colonies don't collide with each other; they may, and do, overlap. Collisions aren't based on actual data and don't represent historical events—that would be an entirely different story. The shape variations imply that the visual display of land area isn't precise during short time frames. Nevertheless, we constrain these variations to maintain accurate visual perception while adding expressive depth. We can adjust several other system parameters to attain different expressiveness.

### Temporal Manipulations

The stories for the fabula of the empires' decline are conducted over a continuous timeline spanning from 1770 to 2010. The event model implements a nonlinear timeline by manipulating the fabula's pace and chronology. The time advances one year per second, speeding up three times in the period when no colonies become independent. These changes in pace let us represent the most dramatic moments—when the colonies gain independence—in detail, while fast-forwarding the monotonous time periods, thus compressing the story's overall time span. This lets us portray more than two centuries of history in less than three minutes, with time rates that are appropriate for visual perception.

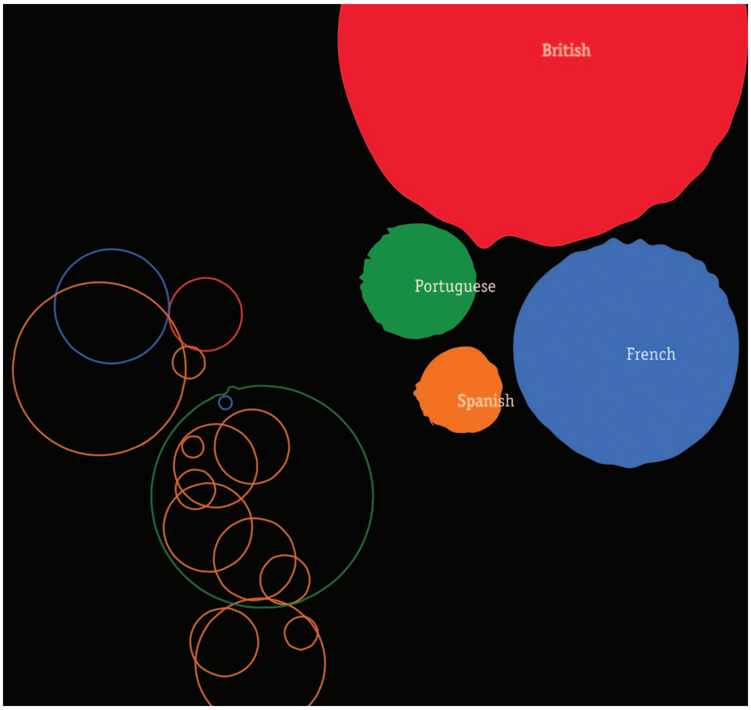


**Figure 3. The British hegemony and the newly independent South America in 1891. Each empire and independent territory is a circle whose area is proportional to that entity's land area. Former colonies are unfilled circles with rims in the corresponding empire's color.**

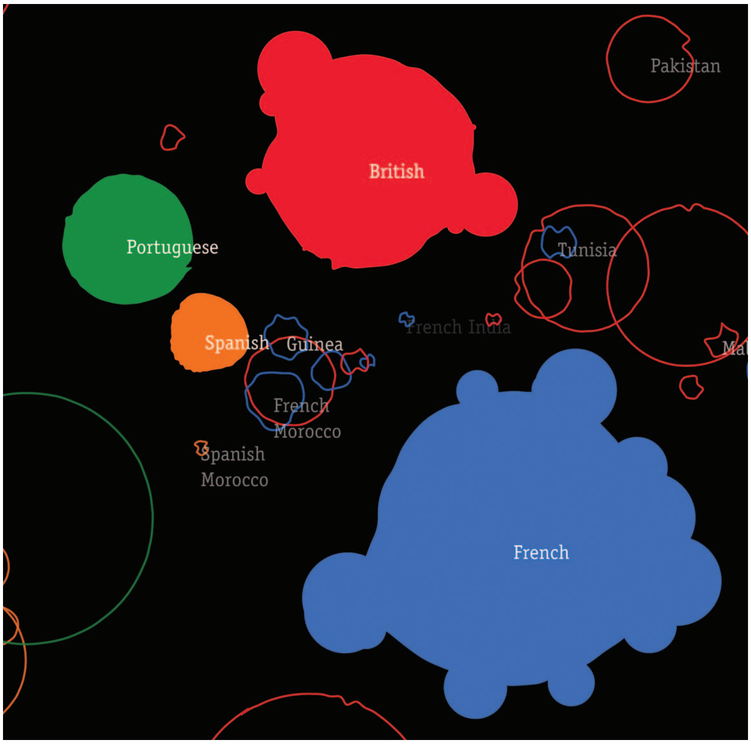


**Figure 4. The moment before the independence of the French African colonies (1960). As the new nations grow, the empire deforms.**

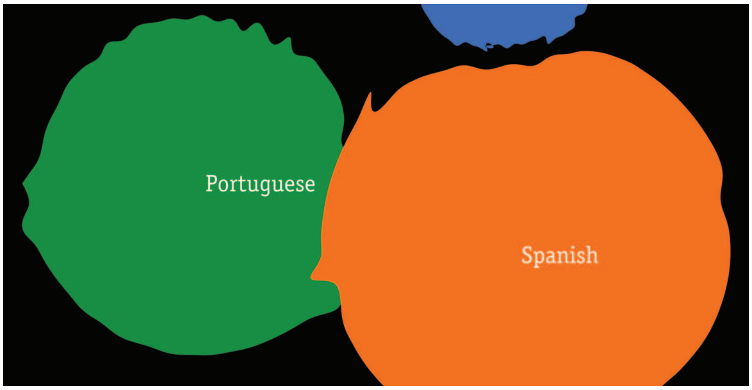


**Figure 5. Soft-body properties—"jigglier" deformations. Empires can collide with each other and display moderate shape variations. This behavior conveys the volatility and instability of the empires' evolution.**

An additional chronological manipulation is the use of subtle prolepses: a growing circle on an empire's rim represents an upcoming independence event. This increases a division's dramatic effect by symbolizing the growing struggle for independence.

### *Implementation*

By using a physics system for the action model, we simulate a set of derivative behaviors stimulated by data. We implement each soft body's behavior by building a particle system of a certain density on the corresponding circle's rim. We then interconnect each particle with springs (creating a skeleton) to attain a soft-body behavior. To implement the springs, we use the toxiclibs 2D physics engine (http://hg.postspectacular.com/toxiclibs). Springs also implement the forces acting in the simulation world, delivering collisions and attractions. To create the traveling action, a spring attracts each body to its geographical position by projecting a typical Cartesian world map in the simulation canvas.

To implement collisions, we dynamically create springs between particles of different bodies when the distance between these particles becomes less than a predetermined value. If the particles are sufficiently apart, we delete the springs. As we previously mentioned, collisions exist between empires and between empires and colonies, but not between colonies (see Figure 6); this facilitates the traveling action.

We implement the division action by creating a body over the corresponding empire's rim. In spite of this condition, the new nation's exact location is determined only at simulation time and is nondeterministic. This small nondeterministic behavior propagates throughout the simulation, exhibiting another feature of generative storytelling.

### *Results*

We're interested in studying how adjusting a model's parameters can achieve different expressions for a fabula (for examples, see http://vimeo.com/6437816 and http://vimeo.com/11506746). By increasing the strength of the springs constituting the circles' skeletons, we approximate rigid-body behavior. The representation becomes truer to the data; however, it loses the expression of fluidity and dissolution of the empires and gains a stiffer personality (see Figure 7). This trade-off would probably be adequate if we wanted to glorify each empire as a relentless entity, but we find this portrayal inadequate to depict the idea of decline.

We can change the collision springs' strength without causing deviant behavior in our stories. By increasing their strength, we augment the repulsion forces. This produces more frenetic behavior, in which the collisions between bodies are abrupt and the new nations are aggressively projected.

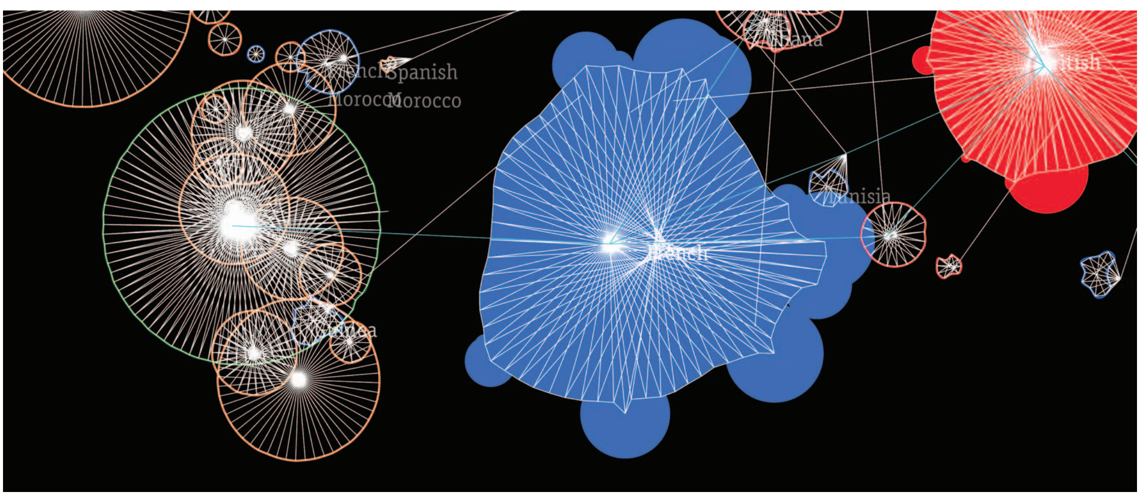


**Figure 6. The springs responsible for the system's behavior. The white lines represent the springs forming each body's skeleton and each body's geographic attractor. The temporary springs that implement collisions between bodies are cyan.**

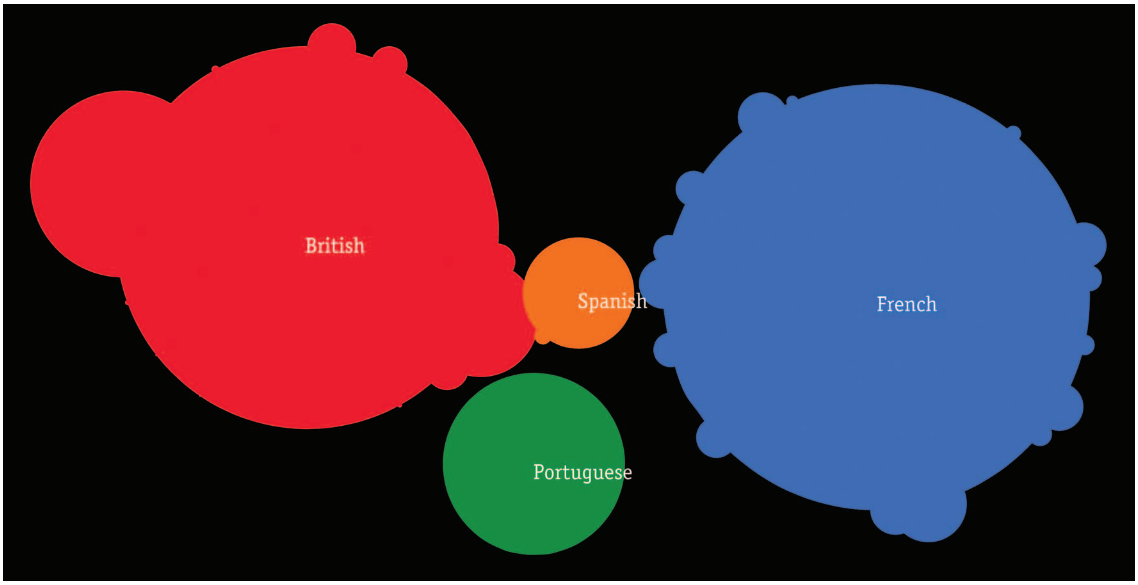


**Figure 7. Increasing the strength of the springs constituting the circles' skeletons to approximate rigid-body behavior. The representation becomes truer to the data; however, it loses the expression of fluidity and dissolution of the empires and gains a stiffer personality.**

This behavior makes the accurate perception of events more difficult, owing to the bodies' greater average velocity. However, we believe it adds a set of entertaining interactions and enhances the ludic experience. In comparison, the story attained by using weaker repulsion forces assumes a soberer, contemplative tone.

By controlling the narrative's pace, we can change the story's emphasis. For instance, in our implementation, the 19th century takes roughly only one-third of the simulation's total time. Moreover, if we use a linear timeline, the empires' stability during most of the 19th century contrasts intensely with their instability and quick land loss during the second part of the 20th century. Conversely, we could slow the narrative even further during highly convoluted periods—for example, independence of the French African colonies. The nondeterminism of the new nations' exact position lets us display a set of story-specific behaviors. Even with only subtle variations in parameters, the interactions between bodies vary from simulation to simulation. (For example, in some simulations, a body might have to "struggle" to reach its geographical position because it was created in the opposite direction. In others, it might find a easy path with no obstacles.) A beauty of generative storytelling is also the ability to generate stories with unique behaviors from the same models and parameters, while adequately portraying the fabula.

On the basis of our experience with the declining-empires visualization, we believe that the ability to generate various stories from the same dataset, the stories' ludic nature, and their inherent expressiveness make them an effective vehicle for conveying vast amounts of information.

We're interested in further exploring the typical parameters for an action model (the adaptive threshold, degree of fidelity, and convergence rate). We can then use these parameters to filter complex fabulas or extract relevant patterns through behavior derived from the models, rather than with data mining. This approach's potential advantage over classic data mining is that we deal with the representation layer sooner and therefore can directly work with our data in a perceptual space. This promotes the discovery of adequate compromises between precise data mappings and loose degrees of fidelity to enhance pattern perception. The story's beauty and identity arise from this compromise.

## References

1. M. Bal, *Narratology: Introduction to the Theory of Narrative*, Univ. of Toronto Press, 1985.
2. P. Cruz and P. Machado, "Visualizing Empires Decline," *ACM Siggraph 2010 Posters*, ACM Press, 2010, article 154.

***Pedro Cruz*** *is an assistant lecturer in the University of Coimbra's Department of Informatics Engineering and a researcher at the university's Centre for Informatics and Systems. Contact him at pmcruz@dei.uc.pt.*

***Penousal Machado*** *is assistant professor in the University of Coimbra's Department of Informatics Engineering and a researcher the university's Centre for Informatics and Systems. Contact him at machado@dei.uc.pt.*