Modifying Entity Relationship Models for Collaborative Fiction Planning and its Impact on Potential Authors*

Alan Tapscott, Joaquim Colàs, Ayman Moghnieh, and Josep Blat

Universitat Pompeu Fabra
C/Tanger 122-140, Barcelona, Spain
`{joaquim.colas, alan.tapscott, ayman.moghnieh, josep.blat} @upf.edu`

Abstract

We propose a modified Entity Relationship (E-R) model, traditionally used for software engineering, to structure, store and share plot data. The flexibility of E-R modelling has been demonstrated by its decades of usage in a wide variety of situations. The success of the E-R model suggests that it could be useful for collaborating fiction authors, adding a certain degree of computational power to their process. We changed the E-R model syntax to better suit the story plans, switching the emphasis from generic types to instanced story entities, but preserving relationships and attributes. We conducted a small-scale basic experiment to study the impact of using our modified E-R model on authors when understanding and contributing into a pre-existing fiction story plan. The results analysis revealed that the E-R model supports authors as effectively as written text in reading comprehension, memory, and contributing. In addition, the results show that, when combined together, the written text and the E-R model help participants achieve better comprehension – always within the frame of our experiment. We discuss potential applications of these findings.

1998 ACM Subject Classification H.1.1 Systems and Information Theory, H.1.2 User/Machine Systems, H.5.1 Multimedia Information Systems, H.5.2 User Interfaces, H.5.4 Hypertext/Hypermedia, J.5 Arts and humanities.

Keywords and phrases storytelling, story planning, Entity Relationship Model

Digital Object Identifier 10.4230/OASIcs.CMN.2014.209

1 Introduction

There have been many attempts to provide computational models for narrative and storytelling, pioneered by Propp's morphology of the folk tale [16]. Narrative models adequate for collaborative fiction planning should deal with several aspects. First, different kinds or genres of narrative need different types of rules, particularly, fiction draws strongly from the authors' creativity. Second, stories should be innovative and original. Computational models for stories often obstruct the creative development of the collaborating authors' contributions [17]. In this paper we introduce a narrative model flexible enough to support a wide variety of fiction stories while laying a strong foundation for all sorts of contributions supporting their internal coherence. Our proposed model is based on the Entity-Relationship (E-R) model, a well-established semantic data representation for database design by Chen[2], widely used in software engineering. It also draws part of its inspiration on Lehnert [10]

* This work has been partially funded by the Spanish Ministry of Science and Innovation (TIN2011-28308-C03-03).

5th Workshop on Computational Models of Narrative (CMN'14).
Editors: Mark A. Finlayson, Jan Christoph Meister, and Emile G. Bruneau; pp. 209–221
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

high level analyses of stories in terms of plot units as arcs in a graph that encodes the plot of the story. The modified E-R model exploits the analogy between the requirements of an information system and the plans for a story. We introduce two key modifications; unstructured object representation and dynamic modelling. We also remove data abstractions and focus on instanced data. The proposed model should support authors' communication for collaborative fiction writing (CFW). As a first step we are interested in testing its impact on potential authors to determine if the modified E-R model has potential as a collaborative story planning tool.

In what follows, first we discuss more in detail our approach within supporting CFW plans. We argue the election of E-R amongst other popular semantic models, and the modifications performed to this model to make it suitable for storytelling. We then discuss the results of a small-scale basic experiment and their implications for CFW. A brief proposal for future work ends the paper.

2 A strategy to support authors in story planning

Within CFW, facts need to be communicated, coordinated and negotiated [12] amongst the writing team. On the other hand, and referring to CFW [9] states that "good extended story telling is constrained by the need to maintain consistency and coherence", as otherwise, poor consistency or coherence can easily lead to losing the suspension of disbelief which is generally accepted as an essential trait of successful fiction. Our own previous work on a *Story on a Wall* [4], a public shared board for collaborative stories, revealed three key factors for the participants/would-be authors: the need of understanding the structure of the story, a high concern for preserving the story consistency when contributing, and a generalized interest on keeping canonicity. Likarish *et al.* [11] state that the use of "a suite of authorship tools that provides quick access to pertinent details would be of immense value". Our next step was to experiment with a digital tool to create and explore multi-authored fairy-tales, *CrossTale*. *CrossTale* contained a rich interface to explore and create new scenes and an underlying (hidden) formal model that set the rules to preserve consistency in the authors' contributions. The results of the experiment [17] pointed towards the need to make more visible the hidden model, as the formal constraints imposed on authors would interfere with the CFW. A good tool for supporting CFW could be a model to plan the content of the story telling (a shared universe SU, in our terminology), visible to the authors, flexible, easy to understand while supporting communication, where consistency could be preserved both by authors and the formal model. We formalise further the desired characteristics of the model:

1. *Flexible symbol representation:* A symbol representation that creative fiction can adapt to the story instead of the story being constrained to the model; authors could change the rules or create new, more suitable ones for the story.
2. *Support for conceptually abstract symbol representation:* The model should be flexible to support both concrete elements of the SU such as characters or locations and more abstract ones, such as feelings or knowledge, and authors should be able to decide which items are elements or descriptors of other ones. This is not the same as abstract data representation, a common feature in computational models, as discussed later.
3. *Informative entity relationship network:* The model should allow authors to express relations between entities in an informative way, possibly through the usage of explicit predicates.
4. *Focus on instantiated data:* Authors would rather use specific characters, such as 'Bob and Larry' and 'how Bob relates to Larry' than 'two instances of character with name

attributes Bob and Larry' and 'how characters relate between them', respectively. Data abstractions could make the model more complex for authors. Next we discuss which model or modified model would fulfil these characteristics to support CFW.

3 Semantic models compared in terms of CFW support

Peckam and Maryanski [15] claim that the benefits of Semantic Models used in Computer Science are *Economy of Expression* (generically useful), *Integrity Maintenance* (very important for consistency and generating suspension of disbelief), *Modelling Flexibility* (whose importance has been indicated) and *Modelling Efficiency* and provided an extensive comparison amongst models commonly used. The capability to establish user-defined logic is positive with respect to *flexible symbol representation*, and allowing users to characterize the relationships and possibly represent them as separate entities is positive for expressing the *relationships network*; hierarchical structures would make the representations unnecessary complex, and there seem to be little advantages for authors in data abstraction, derivation and inheritance provided by entities, thus our focus on *instanced data*.

On this basis, Table 1 maps [15] comparisons to the characteristics introduced in the previous section, which are column headings, with color labels denoting the fitness for the characteristics: green, yellow, and red denote good, medium, and bad fit.

Thus, the Entity-Relationship (E-R) model is the most suited for the characteristics of CFW, and it could even be improved by introducing dynamic modeling (for better flexibility in symbol representation) and unstructured object representations (for enhancing the support to conceptually abstract symbols). The next section presents the E-R model as we modify it to support better CFW.

4 The E-R model modified to enable it for story planning

The E-R model [2] introduced comparatively long ago is still widely used by engineers to design data structures holding real-world input. It is necessarily flexible as its representation should cater for any kind of quantitative data set, regardless of its anatomy, as well as it should address any scenario.

More precisely system architects gather a so called requirements list for an information system and translate it into an E-R model through a process called data modelling; the output fits each requirement within a globally coherent system formally formulated. Our approach intends to exploit the analogy of 'requirements formulated as needs of users', and 'story plans of the authors' both expressed in plain structured English. We suggest that authors develop both the written story plan (in plain sentences) and the E-R diagram simultaneously, and maintain it reflecting the development of the story plan.

The E-R formulation we propose uses the elements of Chen's original model but with different meaning, as software engineering and story planning have different goals. We also attempt to introduce some of the desired characteristics resulting from previous analysis into the E-R model.

Entities represent the *agents* of the story. Any item with any degree of conceptual abstraction could fall into this category. In information systems entities usually denote classes or types, such as animal races or vehicle models. Stories deal with *specific* characters and thus we switch the focus from data *classes* to data *instances*. Instead of dealing with the generic class character, we'll be a character instantiated type many times, identified by some

Table 1 Computer Science Semantic Models compared in terms of CFW.

	Conceptually abstract symbol representation	Flexible symbol representation	Focus on instantiated data	Informative entity relationships network
Entity-relationship model	Structured Object Representation	User specific insertion/deletion constraints. No Dynamic Modeling	Low data abstraction supported. No derivation or inheritance supported	Network of user selectable relations independently represented
TAXIS	Structured Object Representation	User specific insertion/deletion constraints. Transaction and object-oriented Dynamic Modeling	Medium data abstraction supported. Inheritance supported	Strong hierarchical predefined relations represented as entity classes
SDM	Structured Object Representation	Automatic insertion/deletion constraints. No Dynamic Modeling.	High data abstraction supported. Derivation and inheritance supported	Generally hierarchical user defined relations represented as entity classes and independently
Functional	Structured Object Representation	User specific insertion/deletion constraints. No Dynamic Modeling	Medium data abstraction supported. Derivation and inheritance supported	User defined relations represented as functions (no support for hierarchies or networks)
RM/T	Structured Object Representation	Automatic and user specific insertion/deletion constraints. No Dynamic Modeling	High data abstraction supported. Inheritance supported	Generally hierarchical predefined relations represented independently
SAM*	Unstructured Object Representation	Automatic insertion/deletion constraints. Object-oriented Dynamic Modeling	High data abstraction supported. Inheritance supported	Network predefined relations represented independently
Event	Structured Object Representation	Automatic insertion/deletion constraints. Transaction dynamic Modeling	Medium data abstraction supported. No derivation or inheritance supported	Hierarchical predefined relations represented as attributes
SHM+	Structured Object Representation	Built in insertion/deletion constraints. Transaction dynamic Modeling	Medium data abstraction supported. Inheritance over generalization and association hierarchies	Generally hierarchical predefined relations represented as attributes and entities

attribute such as its name. Data abstractions such as generalizations or grouping are removed in order to focus on the instanced level of data. We are no longer dealing with *Characters* in this approach, instead we model Mike and the Butcher.

Relationships represent links between entities, for instance, informing of a fact, such as a contract of marriage between two characters. Since most story entities are instances, relationship cardinality is removed. If Mike and the Butcher have a relationship of friendship, it means implicitly that there is just one Mike and one Butcher. Adding a predicate (such as *marriage* or *friendship*) to the relationship is important to state clearly its meaning.

Attributes provide additional information regarding an entity or relationship. The common E-R formulation uses labels and values, but stories often provide little labels and only values, and entities rarely have any attributes in common. Thus, we avoid labels and store attributes as values. For instance, instead of having a personality attribute with *kind* as its value, a

Figure 1 Example of an information system E-R modelling.

character might have the attribute *kind*. This is more straightforward but less standardized in software engineering. This is not supported in the original E-R formulation, but a certain amount of unstructured object representation is beneficial for SUs.

The following example illustrates the differences between the information system and story planning modelling. *Employees have a Name and ID number. Every Employee has a Payroll assigned. Payrolls have a Gross income value and a Tax deduction value.* This might be modelled by an E-R diagram such as depicted in Figure1.

In a story plan, it is more likely to find a statement such as: *Mike is an unhappy employee with a poor payroll*, which could be represented through our modified E-R diagram (see Figure 2).

Figure 2 Example of a story planning E-R modelling.

Chen proposed a set of rules to translate system requirements formulated as English sentences into E-R diagrams [3], which can be used to translate explicit sentences from a story plan into its E-R model. Specifically the first four rules are simple and easy to use in the context of narratives. They convert *common nouns* into entity types, *transitive verbs* into relationship types, *adjectives* into entity attributes and *adverbs* into relationship attributes. The tenth rule proposed by Chen (meant to convert *clause sentences* into a group of interconnected sub-entities) can help in organizing nested plot data. We propose following a three-step strategy:

1. Formulate the story plan in plain explicit sentences; the narrative plan will be made of "story requirements".
2. Translate the sentences into an E-R model using Chen's rules [3].
3. Merge the E-R models and disambiguate any conflicts.

The merging process involves combining the new information with the one already modelled, and disambiguating any potential contradiction. It involves understanding the new entities and establishing their relationships to existing ones. It is a process that can be almost impossible to automatize or assist due to its subjective nature. For instance sometimes an entity must be transformed into another one, sometimes entities are duplicated or even merged. An author with a good conception of a story plan can perform such task (maybe even refining the concept). The ability to introduce insertion and deletion formal constraints could assist this process. This methodology might be beneficial to planning processes that involve more than one author, especially in fiction genres. Also this methodology could assist non-expert users in using an E-R model.

5 A small-scale basic experiment

The E-R model makes visible for authors the underlying formal model hidden in our previous experiments. Before using the E-R in a software prototype it was necessary to test some basic parameters of collaboration. We also measured contribution to attempt to triangulate with our previous results. The basic parameters were related to cognitive processes supporting collaboration: individual comprehension of a story potentially written by another author, and its recall. If comprehension and memory using E-R were degraded with respect to basic text, the model would be of little practical value. A secondary concern would be the time used by subjects to understand – again, if a much longer time was needed with E-R models, its potential would be minimal.

As a first step we compared the basic individual performance in reading, and contributing to, a story with or without using E-R diagrams. We used the first part of the *Stagecoach* movie synopsis (taken from the Spanish Wikipedia [18]), which seemed a rich enough but short story plan. A volunteer Computer Science graduate created the E-R model from this synopsis. 35 subjects (ages 20 to 65), who signed an informed consent form, were divided into three groups: experimental group 1 had both the text to read and the E-R model, experimental group 2 had only the E-R model, and the control group had only the text. Each group had approximately the same proportion of subjects with previous knowledge of E-R modelling (around 37%). Each subject received a brief training providing a basic understanding of our modified E-R modelling.

Every subject received the corresponding printed material and was given the briefing: *This is an incomplete story plan. Please read it.* This phase lasted as much as the subject felt necessary. Then we measured **comprehension** with a short questionnaire composed by open questions. The same group of judges who selected the questions (based on consensus about their usefulness to determine comprehension) was used to evaluate the answers, and we used free-marginal Kappa coefficients to determine the agreement among them. **Memory** was evaluated by removing the written material and asking subjects to answer a true-or-false questionnaire. Then we evaluated **contribution**: we returned the materials to each subject and gave them the following briefing: *We would like you to contribute to this story plan in any way you want to.* They were free to contribute as much as they wished, at any part of the original text or E-R diagram. **Time** spent in the different phases of the experiment was measured as an indicator of the efficiency of the model.

6 Result Discussion

Our experiment intended to perform a first assessment of the effect of introducing E-R modelling to support planning in collaborative fiction writing. Four main aspects were measured *Comprehension*, *Memory*, *Time* and *Contribution*, whose relevance with respect to CFW we discuss along with the results.

Comprehension. The ANOVA test reveals significant difference between the three group means with regards to comprehension ($p = 0.0132$ $F(2,30)=5.0075$). The post-hoc t-tests revealed no significant difference between the two groups with either text or E-R, but a very significant difference between each and the group with both. Enhanced comprehension likely means more *effective communication* amongst writers, which, along better *coordination* are according to Lowry *et al.* [12] two of the most fundamental processes of collaborative writing. The increased comprehension could be through reinforcement of dual cognitive channels, as proposed by Mayer in his Cognitive Theory of Multimedia Learning [13] (CTML). Another possible explanation to the increased comprehension might lie in the existence of the induced paths Corman *et al.* found in their work related to graphs [5]. In this sense, any graphical representation might achieve this positive result. The positive results together with the theoretical grounding encourage us to continue further. A substantial part (26%) of the group who only had the E-R diagram complained about the lack of a reading order – although we did not see its impact in the measures we took. However, this points at a limitation of E-R models, which usually represent *snapshots* of a data set, when transformations are a fundamental part of stories; we already identified this requirement in a previous section. While representing transformations is somewhat opposed to the nature of standard E-R models, using multiple E-R models, perhaps one per chapter, episode or page, could provide an answer to this problem. Another alternative would be a viable syntax to map the transformations and story progression into one single E-R model, such as the one proposed by Klopprogge [8]. 21% of the subjects provided with the E-R model complained about the confusing syntax of relationships; the roles in relationships were removed (e.g. which character hates and which one is hated in a hate relationship), and this makes a story harder to understand. We did not anticipate this, and using Chen's original relationship role labelling or Corman *et al.*'s approach [5], introducing directed networks of relations, would make the E-R model more understandable for story planning, without compromising simplicity and flexibility. Some E-R experts fixed the diagrams, and some people fixed the text. This might be an indicator of users' motivation for quality / consistency [9]), and is consistent with our previous experiments with Story on a Wall and CrossTale; but it might be due to other factors, such as professional rigor. While the pre-existing knowledge on E-R models surely impacts on the results, the proportion of experts in each group was balanced, so that the differential results would be valid, within the confidence ranges allowed by the quite small numbers of the experiment.

Memory. Despite differences in comprehension, there were not significant differences amongst the groups regarding memory (ANOVA p=0.9341 F(2,30)=0.0682). Knowledge retention could boost the coordination between authors, and Nesbit & Adesope [14] registered its increase through the use of concept maps, which are similar to semantic networks. We need to address this issue in the context of more realistic, long time, conditions of CFW.

Time. Time results are difficult to interpret within a creative context. At a basic level shorter time might be an indicator of some alternatives being more or less efficient than other ones. The average time duration for the training phase was of 2'36" (sd= 28"), and without significant difference between the two groups with E-R. There was a significant difference in reading time among the groups (ANOVA p=0.0016 F(2,27)=8.2061). A post-hoc analysis using t-tests in pairs revealed no significant difference between the groups only text and only E-R, but each of them had significant difference with the group using both materials, which took longer. Viewing the results together, there are no differences in comprehension, memory and reading time for the groups with only text and only E-R, which seems to point towards E-R being a viable alternative to text. The increase in reading time of the group with E-R and text resulted in increased comprehension: as the important point for CFW at a basic level is comprehension, this points towards using both text and E-R for future work. While there were differences among groups, the correlations of comprehension and memory with reading time for individuals were not significant.

Contributions. 6% of the subjects chose to make no contributions There were not significant differences among the three groups in contribution time (ANOVA p=0.9346 F(2,18)=0.0677—2 or 3 subjects per group removed as they took much longer). There was a very weak positive correlation (coefficients ranging from 0.36 and 0.5) between reading time and 1) *word count* for text contributions, 2) *new entities* for E-R contributions, 3) *new but related entities* for E-R contributions and 4) *total new characters* introduced for all the contributions. All the individuals who received only text contributed using text. From the individuals who received only the E-R model, 23% contributed using text and 69% used an E-R diagram—a much higher percentage than experts, around 30%. The individuals in the group that received both E-R and text contributed equally along formats (30% used text, 30% used E-R and 40% used both formats). There is a strong positive correlation between *text contribution word count* and *E-R contribution attributes and relations introduced* for subjects who contributed both text and E-R diagrams (coefficients around 0.8), broadly indicating that they contributed in similar proportions in both ways – not privileging one of them. This leads to suspect that they did it in parallel, textual content corresponding with larger E-R diagrams. At the more basic level, E-R models look as efficient as text, and seem simple enough to be used by an important proportion of non-experts after a very short training. The experiment was oriented towards understanding basic issues in preparation of larger experiments where productive creativity can be tested. At this stage, the different models did not look different in supporting creativity. We did not analyse qualitatively the contributions, and this would be an important for further research in CFW.

The results of the experiment indicate that E-R models are not worse than text in terms of comprehension of the story or recall, the basic cognitive processes supporting collaborative authoring. Using both E-R and text improves understanding, but it requires more time. Thus we can confidently proceed with a large scale collaborative authoring experiment based on E-R models supported by a software prototype. In terms of contributions, we could not triangulate with our previous results. However, a significant portion of subjects without expertise in E-R modelling, and with only an extremely short training, spontaneously contributed using E-R– which is also a positive sign of the potential of the approach.

7 Future work

As indicated earlier, support for the temporal dimension seems key because of the transformational nature of stories, and adding roles to the relationships is also necessary. On the other hand, our modified E-R model could be extended to include recent E-R improvements such as the ones proposed by Hartmann *et al.* [6]. A certain degree of semantics and data structures could be introduced to streamline story planning and assist the authors in their task, without compromising the model flexibility and the authors' predominant role. A clear first step is the introduction of insertion and deletion constraints, optional for authors, as used in our previous CrossTale tool [17]. The model introduced could be used to gather more easily data on the author's construction of the story, seen as important by AI practitioners [1]. This could be used to provide authors with tools that predict and support their needs and actions. If a large-scale is achieved, the data gathered through the model could contribute to build from experimental data computational models of the morphology of different genres (our example was based on a western), just like Propp [16] proposed his morphology for traditional folk tales. Genre/writing technique templates could also provide skeleton frames to support authors during inspirational or creative blockings. The story plan used in the experiment was rich enough, but relatively small when compared to a real site of CFW such as [7]. Scalability of the E-R model is one of the issues to be tested in a more realistic CFW experiment or setting. And in this setting exploiting the computational characteristics of the model to support story coherence, as well as quality of the contributions—indicated earlier—will be paramount. Story telling has been recently acknowledged as an important component of visualization [1]. Reciprocally, generating rich visualizations when the plans or stories expressed through E-R models could help authors and, on the other hand, provide researchers with information to prepare enhanced tools based on predictive models, where clustering techniques will probably be used.

The next experiment we are currently preparing is a web-based large-scale longitudinal collaborative fiction writing aimed at producing stories on a shared universe (its basics already developed). Based on this paper results, the story plan tools will be (formal, not only visual) E-R with improved relationships and plain text. Inter-author communication and visualization of complex story plans will be also used. The focus will be on measuring quantitatively and qualitatively the contributions and some specific aspects are author collaboration dynamics, creativity and consistency / coherence monitoring. At its initial stage the experiment might not be using dynamic models yet.

Acknowledgements. We thank Xavier Ruiz-Collantes, Miguel-Angel Carralero, Jose-Emilio Lavilla, Daniel Giralt and Gabriel Cerón for ideas and help for the experiment, as well as the people who volunteered for it.

References

1. Selmer Bringsjord and David Ferrucci. *Artificial intelligence and literary creativity: Inside the mind of brutus, a storytelling machine*. Psychology Press, 1999.
2. Peter Pin-Shan Chen. The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems (TODS)*, 1(1):9–36, 1976.
3. Peter Pin-Shan Chen. English sentence structure and entity-relationship diagrams. *Information Sciences*, 29(2):127–149, 1983.
4. Joaquim Colas, Alan Tapscott, Ayman Moghnieh, and Josep Blat. Shared narratives as a new interactive medium: Crosstale as a prototype for collaborative storytelling. *International Journal On Advances in Telecommunications*, 6(1 and 2):12–23, 2013.
5. Steven R. Corman, B. Hunter Ball, Kimberly M. Talboom, and Gene A. Brewer. Assessing Two-Mode Semantic Network Story Representations Using a False Memory Paradigm. In Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, editors, *2013 Workshop on Computational Models of Narrative*, volume 32 of *OpenAccess Series in Informatics (OASIcs)*, pages 53–61, Dagstuhl, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
6. Sven Hartmann and Sebastian Link. English sentence structures and eer modeling. In *Proceedings of the fourth Asia-Pacific conference on Comceptual modelling-Volume 67*, pages 27–35. Australian Computer Society, Inc., 2007.
7. Jos Kirps. Galaxiki, 2007. [Online; accessed 30-May-2014].
8. Manfred R. Klopprogge. Term: An approach to include time dimension in the entity-relationship model. In *Proceedings of the Second International Conference on the Entity-Relationship Approach to Information Modeling and Analysis*, ER '81, pages 473–508, Amsterdam, The Netherlands, The Netherlands, 1983. North-Holland Publishing Co.

9 Michael Lebowitz. Creating a story-telling universe. Technical report, Department of Computer Science, Columbia University, 1983.

10 Wendy G Lehnert. Plot units and narrative summarization. *Cognitive Science*, 5(4):293–331, 1981.

11 Peter Likarish and Jon Winet. Exquisite corpse 2.0: qualitative analysis of a community-based fiction project. In *Proceedings of the Designing Interactive Systems Conference*, pages 564–567. ACM, 2012.

12 Paul Benjamin Lowry, Aaron Curtis, and Michelle René Lowry. Building a taxonomy and nomenclature of collaborative writing to improve interdisciplinary research and practice. *Journal of Business Communication*, 41(1):66–99, 2004.

13 Richard E Mayer. *Multimedia learning*. Cambridge university press, 2009.

14 John C Nesbit and Olusola O Adesope. Learning with concept and knowledge maps: A meta-analysis. *Review of educational research*, 76(3):413–448, 2006.

15 Joan Peckham and Fred Maryanski. Semantic data models. *ACM Computing Surveys (CSUR)*, 20(3):153–189, 1988.

16 Vladimir Propp. *The Morphology of the Folktale*. University of Texas Press, Austin, 1968.

17 Alan Tapscot, Joaquim Colàs, Ayman Moghnieh, and Josep Blat. Writing Consistent Stories based on Structured Multi-Authored Narrative Spaces. In Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, editors, *2013 Workshop on Computational Models of Narrative*, volume 32 of *OpenAccess Series in Informatics (OASIcs)*, pages 277–292, Dagstuhl, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

18 Wikipedia. La diligencia — Wikipedia, the free encyclopedia, 1939. [Online; accessed 30-May-2014].

A Story seed text

In 1880, a motley group of strangers boards the east-bound stagecoach from Tonto, Arizona Territory to Lordsburg, New Mexico Territory. These travelers are unremarkable and ordinary at first glance. Among them are Dallas, a prostitute who is being driven out of town by the members of the "Law and Order League"; an alcoholic doctor, Doc Boone; pregnant Lucy Mallory, who is traveling to see her cavalry officer husband; and whiskey salesman Samuel Peacock. When the stage driver, Buck, looks for his normal shotgun guard, Marshal Curly Wilcox tells him that the guard has gone searching for fugitive the Ringo Kid. Buck tells Marshal Wilcox that Luke Plummer is in Lordsburg. Knowing that Kid has vowed to avenge the deaths of his father and brother at Plummer's hands, the marshal decides to ride along as guard. As they set out, U.S. cavalry Lieutenant Blanchar informs the group that Geronimo and his Apaches are on the warpath and his small troop will provide an escort until they reach Dry Fork. As they depart, the stagecoach is flagged down to pick up two more passenger, gambler and Southern gentleman Hatfield as well as banker Henry Gatewood, who is absconding with $50,000 embezzled from his bank. Along the way, they come across the Ringo Kid, whose horse became lame and left him afoot. Even though they are friends, Curly has no choice but to take Ringo into custody.

B Story seed E-R diagram

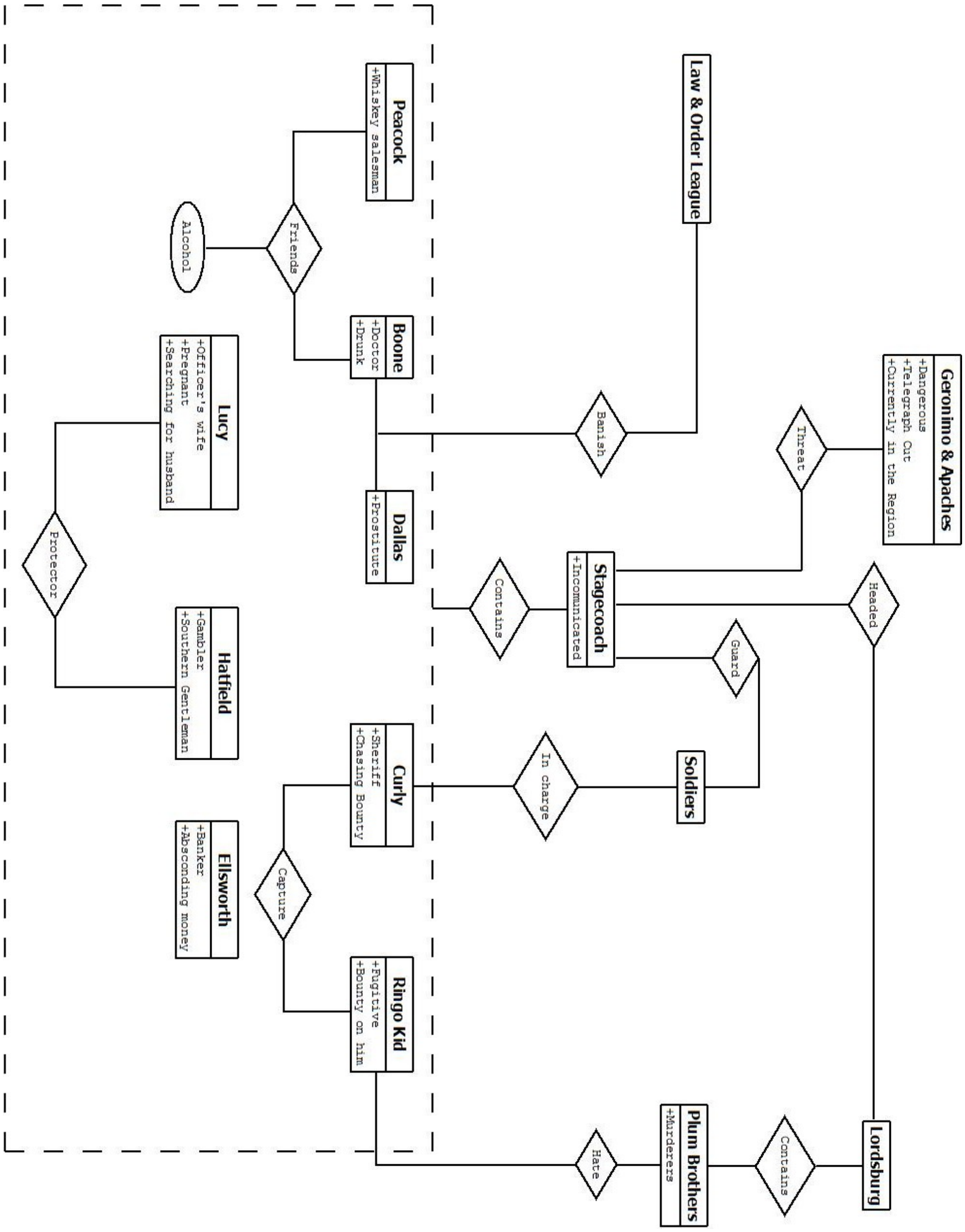

Figure 3 Story seed E-R diagram.

C Result tables

Table 2 Average time spent in the different phases.

	Reading phase	Comprehension test	Memory test	Contribution phase
Text group	124 seconds σ=52.8	261 seconds σ=80.5	70 seconds σ=35.4	288 seconds σ=133.4
E-R group	142 seconds σ=72.4	233 seconds σ=71.6	92 seconds σ=31.2	272 seconds σ=111.7
Text + E-R group	236 seconds σ=69.6	315 seconds σ=107.5	93 seconds σ=34.2	293 seconds σ=92.5

Table 3 Time used in the comprehension phase.

Text vs. E-R	Text vs. Text+E-R	ER vs. Text+E-R
p=0.5170 (no significance) t=-0.6625	p=0.0011 (very strong significance) t=-3.9894	p=0.0101 (strong significance) t=-2.8756

Table 4 Memory and comprehension test averages.

	Comprehension test (0 to 3 points)	(Memory test (0 to 9 points)
Text group	2.325 points σ=0.329	7.417 points σ=1.443
E-R group	2.254 points σ=0.438	7.384 points σ=1.387
Text + E-R group	2.680 points σ=0,167	7.5 points σ=1.354

Table 5 Comprehension test judge agreement.

Average item-total rating correlation	Overall agreement Po %	Fixed-marginal kappa	Free-marginal kappa
0.8900	0.6647	0.3602	0.553

Overall Result: *Moderate agreement*

Table 6 Comprehension test t-tests.

Text vs. E-R	Text vs. Text+E-R	ER vs. Text+E-R
p=0.6124 (no significance) t=0.5150	p=0.0046 (strong significance) t=-3.259	p=0.0088 (strong significance) t=-3.0763

Table 7 Average contribution per group.

Text contributions	**Word count**	**Sentence count**		
Text group	74.6	4.6		
E-R group	56	2.6		
Both group	80.7	4.3		
E-R contributions	**Entities introduced**	**Attributes introduced**	**Old entities related**	**New entities related**
Text group				
E-R group	0.8	1.9	4.1	1.1
Both group	0.5	0.9	5.6	1.6
Overall contributions	**Old characters used**	**New characters introduced**		
Text group	5.2	0.3		
E-R group	3.4	0.5		
Both group	5.5	0.5		