


Please cite the Published Version

Jackson, David and Latham, Annabel  (2022) Talk to the Ghost: The Storybox Methodology for Faster Development of Storytelling Chatbots. *Expert Systems with Applications*, 190. p. 116223. ISSN 0957-4174

DOI: <https://doi.org/10.1016/j.eswa.2021.116223>

Publisher: Elsevier BV

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/628740/>

Usage rights:  Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Additional Information: This is an Author Accepted Manuscript of an article published in *Expert Systems with Applications*.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Talk to the Ghost: The Storybox Methodology for Faster Development of Storytelling Chatbots

Dr David Jackson^a and Dr Annabel Latham^b

^a (corresponding author) d.j.jackson@mmu.ac.uk; School of Digital Arts Research, Manchester Metropolitan University, Righton Building, Manchester M15 6BG

^b (corresponding author) a.latham@mmu.ac.uk; Department of Computing and Mathematics, Manchester Metropolitan University, John Dalton Building, Manchester M15 6BH

Talk to the Ghost: The Storybox Methodology for Faster Development of Storytelling Chatbots

Dr David Jackson^a and Dr Annabel Latham^b

^a (corresponding author) d.j.jackson@mmu.ac.uk; School of Digital Arts, Manchester Metropolitan University, Righton Building, Manchester M15 6BG

^b (corresponding author) a.latham@mmu.ac.uk; Department of Computing and Mathematics, Manchester Metropolitan University, John Dalton Building, Manchester M15 6BH

Abstract

This paper presents the Storybox Methodology which combines a novel framework for structuring knowledge and conversations around a story (D-PAF), with a live chatroom-based training approach that builds the conversation knowledge base via live chatroom interactions. Chatbots have achieved success as intelligent interfaces in education, health, sales and support, but their move towards mainstream adoption has been hindered by the large amount of development resources required, in terms of data collection, preparation, user testing and technical knowledge. The complexity of the development task often necessitates both a system author and a domain expert working effectively together, adding further complexity and risk. Overcoming these barriers could increase feasibility of chatbots in a range of expert contexts. In education, there are groups of learners who do not enjoy reading and writing. Storytelling chatbots might be able to introduce these groups to enjoyable new ways to read and write, having a beneficial impact on their education and future prospects. This paper proposes the Storybox Methodology for the rapid development of storytelling chatbots. Storybox is evaluated by creating, training and testing 'The Ghost', a chatbot enacting Hamlet's Ghost character from William Shakespeare's dramatic tragedy. The results showed that after a period of live chatbot training of only 25 training conversations, The Ghost was able to conduct convincing conversations with participants.

Keywords chatbot, intelligent interfaces, storytelling, response generation, conversational agent, Human-computer conversation system

1. Introduction

There are many challenges in building goal-oriented conversation agents (CAs) that can hold a conversation whilst directing the conversation towards achieving a goal, such as telling a story (Herbert & Kang, 2018; Crockett et al, 2011). The major challenge for the development of CAs is the resource-intensive nature of building a new chatbot, in terms of skills, time and user testing. However, for goal-oriented CAs the additional challenge of knowledge engineering adds to the complexity of the system, with the requirement to capture and structure the knowledge in a form that enables responses to be contextual yet move the conversation towards the goal. This challenge is particularly evident for storytelling CAs, where the ordering and structure of a story is as key a part as the knowledge imparted.

A CA or chatbot can be broadly defined as 'any software application that engages in a dialog with a human using natural language' (Dale, 2016:813). Chatbots that can tell stories could offer a non-judgemental, responsive and fun way to read, write and learn about a number of narrative-based topics for young learners. Educational theorist Michael Stephen Schiro identifies the educational importance of conversational and oral stories within the classroom as a useful technique for contextualising information allowing both teacher and learner to become 'intellectually, emotionally, and physically involved' in learning objectives (Shiro, 2004:46).

In the context of English language and literature education, there are many cases where the ‘social and collective approaches’ to source texts, such as the plays of Shakespeare, have been shown to be highly effective in increasing students engagement and comprehension (Winston, 2015:42). The ability to talk to a character from a story could provide new opportunities for collective and social interaction with source text. However, there are few storytelling conversational agents and no major platforms or systems for producing storytelling conversational agents at the time of writing.

Storytelling chatbots have potential for impact in several fields, particularly in educational approaches to reading and writing literacy. Self-directed reading and writing have a substantial impact on the development of children and young people. The Reading Agency identified a wide range of benefits for children and young people associated with reading, including ‘knowledge of the self and other people, which can also lead to greater levels of social interaction’ as well as ‘improvements in imagination, focus and flow, relaxation and mood regulation’ (BOP Consulting, 2015:19). Literacy is one of the most significant factors in securing and maintaining employment, however in England, 22% of boys and 13% of girls have negative or very negative attitudes towards reading (McGrane et al., 2017) and the National Literacy Trust found that despite enjoying writing being a key factor in overcoming learning barriers, 31% UK 8-12 year olds do not enjoy it (Clarke 2018). Digital interventions for literacy in the classroom, such as talking book software (Lewin, 2002) have been found to enhance learner outcomes in the area of literacy. Digital storytelling can improve writing skills and visual memory capacity in primary school children (Sarica and Usluel, 2016), and enhance foreign language learning, supporting academic achievement, learning motivation and critical thinking (Yang and Wu, 2012). The ambition of the research presented in this article is to use modern technology to tell stories in a different, interactive way, to encourage an interest in reading and writing in readers for whom standard approaches to literacy do not work.

This paper describes the Storybox architecture and methodology, which has been especially designed to develop storytelling CAs quickly, for use in education and in scenarios where CAs have a linear narrative purpose such as health education, customer service and entertainment. The paper also presents an evaluation of the methodology and architecture through the development of a storytelling chatbot called ‘The Ghost’. Storybox simplifies the challenges faced in development of chatbots and speeds up development by focusing on the unique requirements of storytelling chatbots, as opposed to general purpose chatbot models.

The Storybox methodology includes two key phases. Phase one involves structuring the knowledge of the fiction to be told into short snippets of conversation structured using a novel Digressive-to-Progressive Answering Framework (D-PAF). Phase two involves a method of training a storytelling chatbot using a small number of live storytelling sessions.

The novel contributions in this paper are:

- A methodology called Storybox for developing new storytelling chatbots with the potential to educate learners in English and other humanities learning contexts, as well as in other areas such as health, sales and support.
- A narrative-based framework called D-PAF, for structuring story knowledge for conversational interactions.
- A case study of a chatbot called The Ghost developed following the Storybox methodology, that allows participants to speak to the character of Old Hamlet’s Ghost in Shakespeare’s play Hamlet.
- An evaluation of the Storybox methodology in developing and training The Ghost.

In this paper, Section 2 describes the challenges relating to chatbot creation in the context of storytelling and also outlines procedural aspects of storytelling that relate to the challenge. Section 3 introduces the Storybox framework and section 4 proposes the D-PAF architecture for structuring story knowledge. Section 5 introduces the Storybox architecture, and section 6 describes the Storybox methodology. Section 7 describes an evaluation whereby a chatbot called The Ghost was developed following the storybox methodology and trained and evaluated by participants. Section 8 presents initial results and discussion and section 9 outlines conclusions and future work.

2. Related work

2.1. Conversational agents

Conversational agents (CAs) are software agents that allow people to communicate with computer systems using natural language dialogue (O’Shea, Bandar, & Crockett, 2011). In the field of CA research, there are important distinctions between the terms conversational agent and chatbot, with chatbots defined as systems that can respond to simple queries and that are suited to simple Question and Answering (Q&A) scenarios (O’Shea, Bandar, & Crockett, 2011). CAs, however, can engage in lengthy discussion by replicating human communication. CAs are backed up by an advanced CA engine that uses natural language processing and dialogue management techniques to maintain the context over the course of the conversation, and to enable mixed-initiative dialogue (i.e. users can ask questions without following a predetermined order) (Latham et al, 2012).

Recently, ‘virtual assistants’ (ie. chatbots) have gained popularity as a means of improving customer service and support, as they are able to respond automatically and quickly to common questions, and also offer companies market intelligence through extensive analytics facilities (IBM, 2019; LogMeIn Inc., 2019; Filipczyk et al, 2016). In healthcare, a survey carried out by Montenegro, da Costa and da Rosa (2019) found over 4,000 papers relating to conversational agents in the field and predicted, based on its findings, that the use of agents in the field would grow especially in the area of health education. However, despite their promise as a valuable tool for commercial and societal challenges, the early obstacle of the expensive and time-consuming development of a knowledge base on which to construct conversations is still a considerable limitation of their use (Neumann et al 2019; Filipczyk et al, 2016). Therefore, CAs are rarely found outside of scalable areas such as HEI online learning environments (Tegos & Demetriadis, 2015) or scenarios where the knowledge base of the agent is strictly defined by the nature of the subject matter (e.g. computer science in Latham et al. 2012 and Griol 2016).

To create a new chatbot, a system must be selected or developed that is capable of a number of tasks including:

- sourcing, storing and indexing a large amount of response content in a meaningful way for the task it is intended for (e.g. storytelling);
- providing an interface for interacting with the chatbot;
- interpreting and responding to utterances from a human user in a way that is consistent and meaningful.

The chatbot must then be trained on the specialist content in order to converse with their users (Yan et al 2018). The amount of content considered to be sufficient to train chatbots is non-trivial. Facebook’s PERSONA-CHAT dataset consists of “164,356 utterances between crowdworkers” (Zhang et al., 2018:1). Herbert and Kang observe that this process also typically requires a split between the domain expert and a system author, with “high-level technical or syntactical analysis skills” (Herbert and Kang, 2018:342) to systematise domain expertise for meaningful input and output. This represents additional complexity and resourcing in the production process. McNeal and Newyear (2013) describe the lengthy process of creating chatbots for library services using popular chatbot language AIML, estimating that to create a convincing chatbot, somewhere in the range of 60,000 + categories is required to ensure a correct response. To overcome this challenge, researchers are working on systems that can generate CA responses from text found on social media, Internet searches and unstructured documents (Arsovski et al 2019; Wang et al 2019; Yan et al 2018). Conversation generation is a promising idea for speeding up development of Q&A chatbots, however this does not yet address the problem for goal-oriented conversations, where history and context is critical (De Kleijn et al 2019).

Latham et al (2012) described two main approaches that CAs adopt to understand user input: the semantic-based approach (which analyses language constructs and meaning, or scores the semantic similarity of phrases) and the pattern-matching approach (which, rather than attempting to understand user inputs, uses an algorithm to match phrases to a knowledge base of stimulus patterns). Whilst pattern-matching currently works best for extended dialogues, the development of stimulus-response pairs (known as a CA script) is a skilled and labour-intensive task, requiring the anticipation of conversations

in advance. There follows an extended period of user testing, during which conversations are logged and analysed alongside the CA scripts in order to evaluate and improve them. Adjustments include dealing with conflict resolution (i.e. ensuring that the correct response ‘wins’), the addition of new, unanticipated, patterns and the extension of stimulus patterns to incorporate the different use of language of the target users. Thus, despite the promise of intelligent, natural communication with humans, CA research has yet to satisfy this goal.

2.1.1 CAs in educational systems

Known as *pedagogical conversational agents*, CAs as educational tools have been successful as interventions in a wide range of educational environments, supporting both learner and teacher. Typically, the CA takes on the conversation patterns of human tutors to provide additional conversational learning tools (Graesser & Li 2014). These approaches have been found to be more effective than textbook or other non-interactive learning materials (ibid) and equally as effective as other types of rich media in educational contexts (Sebastian & Richards 2017). In online learning applications CAs have also been shown to be effective at promoting peer-to-peer learning by asking questions that encourage academically productive conversations (Tegos et al 2015; Tegos 2017) significantly improving learning outcomes. Through tailored social intervention in these contexts, CAs have also been shown to create positive environments for group learning (Kumar and Rosé 2014). Coronado et al (2018) supplemented a Q&A chatbot with social dialogue, in what they describe as a cognitive assistant for learning Java programming language.

While CAs have been used to support learning, their use in delivering automated conversational tutoring is less common. With Oscar, Latham et al (2012, 2014) have shown that it is possible to improve learning outcomes in conversational intelligent tutoring systems by automatically profiling a user’s preferred learning style and adapting the tutoring style of the CA accordingly. Graesser et al (2017) explored the use of text-based and embodied CAs for conversational assessment of a learner’s mastery of a subject.

The common factor in the range of pedagogical CAs in current research is that purely because of the nature of a two-way conversation, CAs can offer a more social experience for learners, and learning is an inherently social process (Jones & Issroff, 2005; Wang & Wu, 2008).

2.1.2. CAs in stories

In storytelling, development of CAs has focused on interactive fiction systems. In early interactive fictions (IF) like Adventure, early stage natural language processing models enabled players to interact with a story using simple text prompts (‘go north’, ‘take apple’). Agent-based storytelling experiences such as Lyotard (Loyall & Bates, 1997) and Galetea (Short, 2000) followed that used IF authoring systems to allow users to talk with CAs with limited fluency. These systems required complex rules to model the responses of the characters, and such underlying complexity made the fundamental interactivity of the experiences poor: the user does not properly engage in conversation because they spend the majority of their time finding combinations of words that will spark a response from the agent to move the story forward. Façade by Mateas and Stern (2007) represents one of the most developed forms of digital narrative multi-agent storytelling. It uses an AI that knows “how stories are structured” in order to construct “new story-like experiences in response to the player’s real-time interaction” (Mateas & Stern, 2003:2). The relative success of the system came at the expense of much time and effort from its creators, taking six years (Windrip-Fruin, 2009) to complete. To date there have been no other games based on Façade’s ground-breaking technological approach.

The main use of CAs to tell stories has been in computer games; in this context the player interacts with a non-playable character (NPC). Typically, the type of interaction is limited to multiple-choice style dialogue choices presented to the player and is used within the game as “a method of making gameplay decisions – with different dialogue options altering gameplay”

(Windrip-Fruin, 2009:57). The use of game mechanics to motivate conversation leads to a style of conversation that is extrinsically motivated by the objectives and challenges of the game, to the extent that a player is unlikely to value an otherwise interesting conversation with an NPC that does not further their chances of success within the game. More recently, commercial systems have been developed that allow for the creation of Storytelling CAs (SCAs) as part of digital media projects. Systems such as Charisma.ai (Gadney 2018) are designed to allow drama-based applications to “break the fourth wall” (ibid.:online) and speak to the user, adding new AI features such as emotion sensing to the core conversational experience in an attempt to “increase emotional engagement with the character” (ibid.) of a story.

The use of SCAs in heritage contexts is also increasing. The Forever Project (Ma et al., 2017) allows users to interact with digital recordings of Jewish Holocaust survivors at the National Holocaust Centre, UK by asking questions, via a microphone, to the SCA. A bank of 800 to 1,200 questions and answers allows the SCA to respond to user questions with a good level of fluency. The Forever Project exemplifies how SCA applications can help to connect learners with human subjects and characters in a way that is natural and compelling.

2.2. *Storytelling structures*

2.2.1. *Node based structuring methods*

Node based digital storytelling is a popular approach to adding interactivity. Nodes are designed to marry the top-down planning of the storyteller with the bottom-up input of the audience (Ryan 2006) by introducing points at which the learner can make a choice that affects what they read next. As described by Mateus and Stern (2003), in these structures,

“each node is a finely-crafted chunk of content such as a plot event The player is given the ability to traverse [them], and the resulting sequence of nodes constitutes the experience of the narrative.”

Mateas & Stern, 2003:4

Whilst node based structures take many forms, the majority of interactive storytelling narratives use some variant of a decision tree structure to organise the learner’s path through nodes (Murray 2017, Ryan 2006, Crawford 2004). A popular example of this is the Choose Your Own Adventure series of books in which readers are able to read part of an adventure and then choose what to do next (fight the attacker or flee, for example). Decision trees have been criticised as not offering the reader enough choice, whilst simultaneously being onerous in terms of content production (Crawford 2004). Other forms of node-based structure listed variously by observers such as Marie-Laure Ryan and Janet Murray include *maze* structures that use spatial navigation as an interactive narrative device and *rhizome* or *network* structures that allow the audience to move backwards and forwards, arbitrarily, through the events that make up the story.

Ryan (2006) notes that the purpose of node-based digital storytelling architectures also varies. Some architectures (such as the network structure just described) give learners agency in choosing the different ways to navigate through a predetermined story, whilst others such as the branching narrative represent a pattern of choices that can result in being told different stories (Ryan 2006:102). The Storybox methodology proposed in this paper will use branching methods and the notion of nodes from techniques of interactive storytelling to provide conversational interactivity along a predefined storyline.

2.2.2. *Linear story structuring methods*

The idea that stories conform to certain patterns that can be utilised in procedural models of narrative has a rich tradition. By studying 300 Russian folk tales, Vladimir Propp was able to observe the overall morphology of the literary genre, describing this type of tale “according to its component parts and the relationship of these components to each other” (Propp, 1968:8). Propp was the first to document and

describe how certain types of traditional linear stories share with each other similar events in a particular order. This idea was later popularised in the West by Joseph Campbell (Campbell, 1949) whose work ‘Hero with a Thousand Faces’ put forward the idea of the *monomyth*, or archetypal narrative comprising a structure which can be found in all the major myths. Campbell’s work has been influential in structuring commercially successful narratives such as Star Wars (Murray, 1997:186). In the field of behavioural psychology, Brewer and Lichtenstein (1982) found experimental evidence to support the idea that particular structures of narrative are influential in defining whether we classify a narrative as a story or not. For example, if the narrative employs one of three key structures: surprise, suspense and curiosity and the structure is not completed by the end of the story (if we do not find out who the murderer is in a murder mystery or whether the lovers are reunited in a love story), it is usually considered incomplete or erroneous by its learner. Robert McKee defines story structure in less concrete more absolute terms as a “series of acts” leading to a climax which brings about “absolute and irreversible change” (McKee, 1997:47). Putting the onus on the writer rather than the story, McKee defines the plot of the story as “the correct path” plucked from a “dozen branching opportunities”. All of these story definitions point towards the notion of story structure as one highly regulated by the nuanced perception of its readers, one that is necessarily repetitive in essence and highly sensitive to interruptions and errors in its telling.

In the case of linear storytelling it is essential that the storyteller plays an authoritative role in storytelling experience and maintains control. Whilst the listener has some licence to ask the storyteller questions and to even poke fun at the storyteller’s story there is no sense in which the listener has control over the story outcome as in interactive story texts, or that that such a prospect would even be desirable. As literary critic and novelist Sontag (2005) observed, the role of the storyteller as sole controller of the narrative does not represent tyranny but is part of what we enjoy about stories. What interests us most about fiction is “the idea that events happen in a specific causal order” (Sontag 2005: 11); those of us who read nothing else “will read for plot”. In the context of conversation therefore, it is always the duty of the storytelling agent (human or non-human) to guide the conversation back to the storyline. This notion forms the basis of the Storybox Digressive-to-Progressive Answering Framework (D-PAF) proposed in this paper and described in Section 4.

3. The Storybox Framework – Hamlet’s Ghost chatbot

Storybox is a framework for creating chatbots that can hold conversations and tell stories. It has been designed to allow the chatbot to respond to queries and narrate as one of the story’s characters. The pedagogical aim of chatbots developed using the Storybox methodology is to provide learners with opportunities to explore stories that are relevant to their learning (for example, the two Shakespeare plays that feature in the National Curriculum (Crown, nd.)) and ask questions of one of the characters in the story, developing learners’ understanding of the text whilst also developing their confidence and enjoyment of reading and writing through the interface. The Storybox methodology does not adopt specific learning ontologies but rather trains with target learners to pick up their styles of questioning in a live environment. Storybox has been designed to minimise the amount of pre-defined content that the chatbot needs in order to answer questions correctly. Storybox additionally offers the opportunity for learners to be involved in the content development of the chatbot, increasing their engagement with the story, giving learners an insight into how stories are crafted and providing a viable way of developing the story conversation that does not involve an impractical amount of out of classroom effort to set up.

3.1. Storybox Architecture

The Storybox system architecture for a storytelling chatbot was designed to allow the collaborative training of the chatbot’s conversation with users based on the story script. Using this approach, a corpus of conversation couplet scripts is built for future use by the CA. Different components of the same architecture also allow conversations between the CA and the user, once enough script couplets have been generated.

The Storybox architecture is shown in Fig. 2.

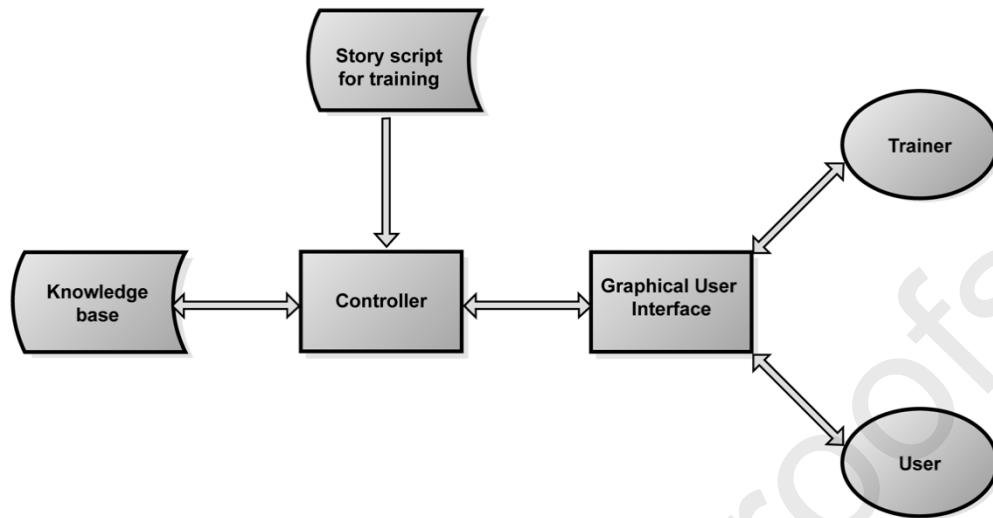


Fig. 2 Storybox System Architecture

In Fig. 2:

- *The user* is the person seeking to interact with a character from the story
- *The trainer* is the human agent who initially answers questions from students instead of the conversational agent storyteller
- *The graphical user interface (GUI)* takes on the form of a chatroom which facilitates a conversation between the trainer and the user or the conversational agent and the user
- *The controller* has two modes: chatbot mode and training mode, which can be switched between via the GUI. In training mode all questions are answered by the trainer using the framework illustrated in Fig. 1 based on a bespoke story script for training. In Chatbot mode, the controller uses the knowledge base to automate story-based conversations with the user.
- *Story script* provides structured content for the Digressive-to-progressive Framework (see Section 4) appropriate to the current character story
- *Knowledge Base* contains the conversational data from training mode interactions for use in chatbot mode.

In a pilot study described in this paper (see Section 6), the Storybox methodology was followed to develop The Ghost from Shakespeare's play Hamlet, reimagined as a chatbot. Hamlet's Father's Ghost chatbot has a broad pedagogical aim of encouraging interaction with the characters and themes of the play through written input and reading the outputs of the ghost. It also encourages reflection on some of the themes of Shakespeare's play Hamlet, a regular text in UK secondary school English Literature at GCSE and A-Level.

The Storybox methodology for creating storytelling chatbots consists of two phases. The first phase of the methodology relates to the design of the story conversation by engineering the knowledge according to a novel generic framework called the Digressive-to-Progressive Answering Framework (D-PAF). The second phase involves one or more live training sessions designed to populate the chatbot's supporting database with question and answer material that will allow it to select appropriate answers. The D-PAF framework will now be described (Section 4), and each phase of the Storybox methodology will be described in Section 6.

4. Structuring the story knowledge - the Digressive-to-Progressive Answering Framework (D-PAF)

In linear storytelling, a story progresses from start to finish via a set of events that happen in a causal order, each moving the plot forward, and this helps to maintain the interest and understanding of the listener. However, one of the content problems for storytelling CAs that was discovered during a preliminary evaluation of the Storybox method is that the story is often not told. Without a framework that forces the story to progress, the subject of the conversation often becomes the character that the user is interacting with. These conversations then follow a series of unpredictable *getting to know you* type questions, (for example *Hello what's your name?*, *Okay, where are you from*, *What do you like to do in London?*). As well as leading to undramatic conversations between the character and the user, this results in little to no similarity between questions or question types that could be easily predicted in advance. Within the Storybox methodology this tendency by the user to ask out of scope questions is described as *digressive*: the question typically digresses to a greater or lesser extent from the strictly linear progression of the story. For example, a content-type rule: e.g. all chatbots must have an answer for what they like to do in London, would not necessarily answer any future questions. Thus providing an answer node to the many hundreds or thousands of other questions like this one represents highly ineffective effort, given that most conversations with the character chatbot would only last for 5-15 questions. The Facebook Chitchat models show that in addition to the high cost (in terms of money, storage and time) of big data approaches to *getting to know you* chatbot interactions, it is still difficult to provide contextually consistent answers (Zhang et al, 2018).

Early prototype testing also provided a potential solution to this content problem. When the story character systematically moved the plot forward during interactions, the user's attention was refocused on a relatively limited and predictable set of curiosities provoked by the story's plot. This is consistent with research by Gerrig, Love and McKoon (2009) that shows that storytelling is at one level a process of deliberately raising questions in the mind of the reader and then tactically answering those questions later on in the plot (Gerrig et al., 2009). The Storybox methodology proposed in this paper uses the idea of raising deliberate questions throughout the story in order to predict and limit the number of responses that the user has about the text. It calls these interventions *Cliffhangers*.

4.1. D-PAF three-part response

When a user asks a question about the story the character is telling them, the force of the question is typically digressive – it seeks to direct the storyteller to a greater or lesser extent away from the standard linear text, whereas the storyteller of a linear story seeks to progress the storyline from its beginning to its end. In order to respond to the learner yet progress the story towards the next event, D-PAF produces a structured, composite answer by the chatbot. The D-PAF structure traverses from digression back to the progression of the story (see Fig. 1, number 1). When the user asks a question, (e.g. 'Where are you now?'), the chatbot's composite response is structured in three segments:

- 1) **Free-text answer**: the first segment is a specific response to the user's question that has been generated on the fly by a writer in conversation: it typically repeats some of the language used by the user, e.g. *How did you get trapped in a computer* is responded to with *What is a 'computer'?*
- 2) **Segway**: the second segment of the same answer has been selected from three or four prepopulated joining texts which sensibly link the subject that the user asked about with the content in the Cliffhanger that will move the conversation on. To prepopulate the Segway, the author of the story must make assumptions about the kind of questions that their user will ask at this stage in the storytelling.
- 3) **Cliffhanger**: the third and final segment of the response is fixed and moves the story on to the next critical stage in the narrative and provides new curiosities for questioning.

Story question reconciliation design

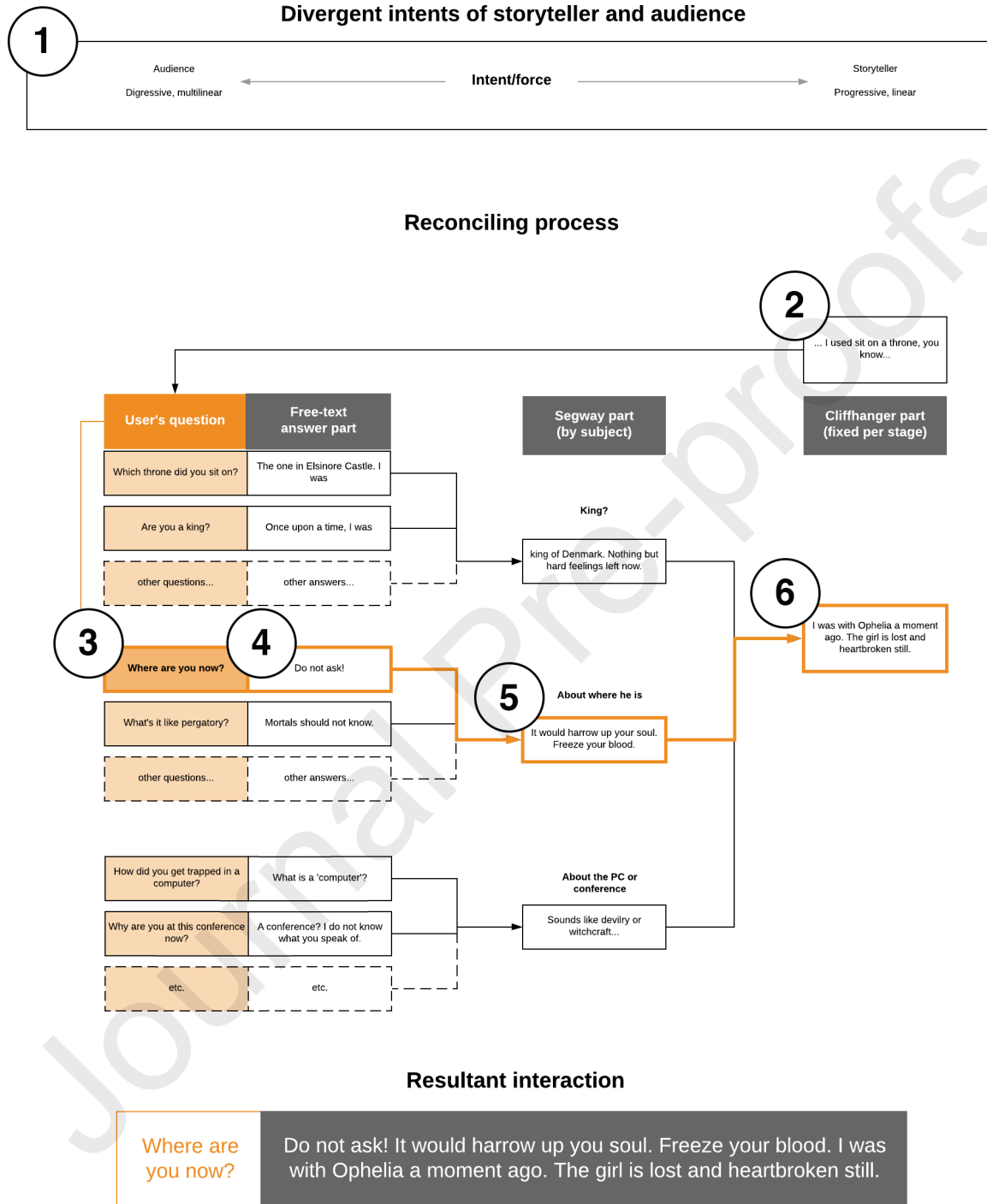


Fig. 1: Digressive to Progressive framework (D-PAF) example from section of Hamlet's Ghost chatbot
 Fig. 1 provides an example of how D-PAF creates composite responses. The example illustrates how dialogue is constructed for the character of The Ghost from the Shakespearean play Hamlet, which was developed during the evaluation of Storybox (described in section 7). The chatbot has just offered some information about its character (2) ('...I used to sit on a throne you know'). In response, the user has asked

a new question (3) ('Where are you now?'). The free-text part of the answer (first segment) is designed to fit appropriately with the user's question (4) ('Do not ask!'). In the Segway part, questions *About where he is* were correctly anticipated as a resultant line of questioning and so the predefined Segway is selected (5) ('It would harrow up your soul. Freeze your blood.'). Finally, the ghost switches subject in the Cliffhanger (third segment) to another character Ophelia in order to provide a new question stimulus (6) ('I was with Ophelia a moment ago. The girl is lost and heartbroken still'). By pre-writing the Segway and the Cliffhanger segments, the scripting allows the writer to fit their answers to the overall direction of the narrative at speed in the live environment.

5. Storybox Methodology

Table 1 shows the Storybox Methodology for developing a new storytelling chatbot. Development of a new Storybox chatbot instance is split into two content development phases, which will now be described.

Table 1. *2-phase Storybox methodology*

<p>Phase 1: Pre-populate storychatbot framework</p> <p>1.1. Design story content for chatbot training</p> <ol style="list-style-type: none"> a. Choose character and story for chatbot b. Identify the key events in chosen story c. Write answers to the main questions raised by each event <p>1.2. Pre-populate Storybox story script trainer framework</p> <ol style="list-style-type: none"> a. Populate cliffhanger segments with material from key events b. Populate segway segments with questions raised <p>Phase 2: Live chatbot training</p> <p>2.1. Set up two-terminal training system</p> <p>2.2. Conduct over 20 human-human conversations</p>

5.1 Phase 1: pre-populate the story chatbot framework

5.1.1. Phase 1.1. - Design story for chatbot training

In order to set up the chatbot for training, educators must first select and plan the key elements of their story and character:

- a. Narrative text is selected for Storybox (such as a novel, play or film) and then a character is chosen from the story to act as narrator. The character chosen may be the protagonist but could also be another character that has a good overview of the events of the story. Ultimately, this is a creative decision.
- b. Up to ten key events are selected that the character will tell the user during their conversational interactions. Useful guidance on approaches to plotline development are found in McKee (1997) and Miller (2004).

5.1.2. Phase 1.2. – Pre-populate Storybox story script trainer framework

Once the planning is complete, the chatbot is pre-populated within Storybox itself. When the text is entered the tone of voice and perspective of the character is considered and the narrative retold from a first-person perspective:

- a. Using the key events identified during the planning stage populate the Storybox system is with up to 10 Cliffhanger texts that are designed to pose specific questions for the user.
- b. Segway sections are written for each of the Cliffhangers. The writer must imagine approximately four possible questions that a user might ask in response to the previous Cliffhanger and write texts that feasibly link to the next Cliffhanger. Segways should also leave a little room for free text at the beginning of the utterance. Each Segway also has a label to allow to quickly identify

the right text from the system's drop down menu when live chatbot training (see Fig. 3). The Segway parts are written in a way that allows scope to write responsive free text comments at the beginning of the chatbot answer (see Fig. 4)

Once the Segway parts have been created for each Cliffhanger they are formatted as a JSON file for use within the Storybox storyteller chatbot training system.

5.2 Phase 2: Live chatbot training

In order to provide more question-specific answers in automated conversations, the chatbot is trained with live conversations between the designated trainer playing the role of the story character (chatbot) and a group of users who ask questions about the character and story, to which the trainer responds. This use of a human operator acting as CA, known as the Wizard-of-Oz technique, has been used in other contexts to improve the flexibility of educational software applications (Ahn et al., 2017). Here the trainer is responsible for providing all response data during the live conversational sessions.

5.2.1 Phase 2.1 Set up two-terminal training system

Two networked computer terminals are set up. One is for the trainer and the other is for the user(s). Both log into the Storybox system chatroom to different profiles: the trainer logs on as the character profile and the user logs on as their own avatar.

5.2.2. Phase 2.2 Conduct over 20 human-human conversations

1. The user writes a greeting or a question and sends it.
2. The trainer responds by first deciding which Segway is best suited to the question (see Fig. 3. for illustration) and then by writing a piece of free text tailoring the answer to the user's message (see Fig. 4).
3. When the trainer sends the response, the Storybox system composites together all three segments, i.e. free text + Segway text + Cliffhanger text as one message (Fig. 5).

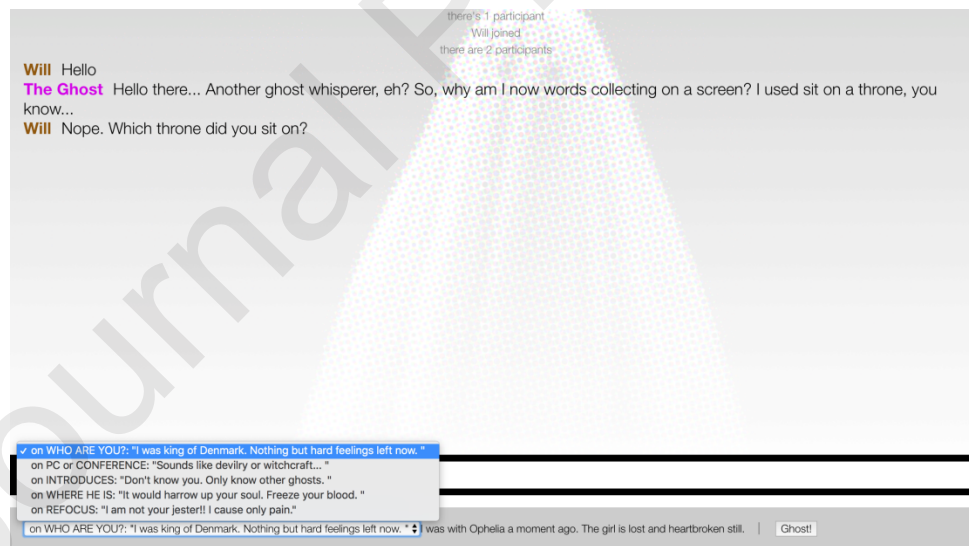


Fig. 3 Trainer selecting Segway text in Storybox graphical user interface



Fig. 4 Trainer completing free text response in Storybox graphical user interface

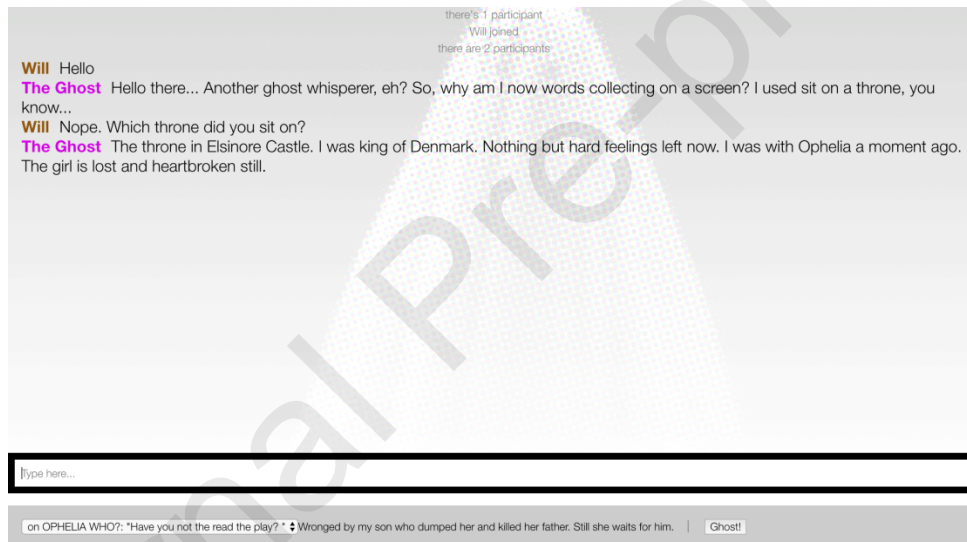


Fig. 5 On hitting Enter, the full text response comprising free text, Segway text and Cliffhanger text ('I was with Ophelia...') is composited and sent to the user

It is important that the trainer and the user take turns so that proper couplets can be formed. This means that the user only ask one question at a time and the trainer only offers one text as an answer. On reaching the end of the story script, the storytelling conversation is finished and both parties can end the session or start a new conversation from the beginning.

The purpose of this disciplined approach to question and answer is the creation of Q and A couplets assigned to specific stages of the plotline. Each of these couplets is stored in a database on a mongoDB document database in the following fashion:

```
{
  "id": 1535549644489,
  "session": "pen-corner-1535549466829",
  "message": "Nope. Which throne did you sit on?",
  "response": "The throne in Elsinore Castle. I was king of Denmark. Nothing but hard
feelings left now. I was with Ophelia a moment ago. The girl is lost and heartbroken still.",
  "stage": 1
}
```


In the database: }

- *Id* is a unique key for the document in the database
- *Session* tracks the particular session of the user with a key shared by all documents in a particular conversation
- *Message* documents the utterance of the user
- *Response* documents the response to the message by the trainer
- *Stage* tracks the stage in the storytelling process that the message and response belong to.

When the Storybox system begins to answer questions automatically these data allow it to poll results based on both question and answer textual similarity, limited by the story stage parameter. It uses a weighted text search, prioritising similarities found in the user question by a factor of 10 in comparison to responses. This allows the system to fall back on potentially relevant answers to the question in the answer text, while preferring similar questions that showed a higher likelihood in informal testing to provide relevant matches.

6. Evaluation of Storybox methodology

In order to evaluate the Storybox methodology, a prototype chatbot called The Ghost, from Shakespeare's play Hamlet, was developed, following the method described in section 6 (Table 1). The purpose of the evaluation was to assess the feasibility and suitability of the Storybox methodology in a specific context, i.e. education, to produce a convincing character-based agent that could tell a story with a relatively small amount of 'Wizard-of-Oz' training. In order to evaluate this most effectively, educational practitioners and researchers were chosen as experts in the field, to participate in the study.

In step 1.1a, the character The Ghost, from Shakespeare's play Hamlet, was chosen in consultation with organisers at the Playful Learning 2018 conference to tell a story that would be relevant to learning contexts for a large cross-section of high school users (11 to 18 years old) studying subject such as English Literature and Drama. The participants at Playful Learning were educators and educational researchers with expert knowledge in the use of digital technology in educational settings at both high school, college and professional settings, but no knowledge of the system or of storytelling chatbots. In step 1.1b, the knowledge in the play was analysed to extract key events and themes. It was decided that just a few elements of the three-act play could be covered, rather than a comprehensive retelling of events in the play (Table 2).

Table 2. Key events and themes for Hamlet's Ghost chatbot to cover, selected by researchers

- | |
|---|
| <ol style="list-style-type: none"> 1. The protagonist was a king 2. He was murdered by his brother 3. Another character called Ophelia dies tragically 4. Ophelia was wronged by young Hamlet who rejected her and murdered her father 5. The king's brother is killed by young Hamlet as an act of vengeance 6. The king's role in the play as an ambiguous agent, either a devil or a wronged soul. |
|---|

Next, in step 1.2, Segway and Cliffhanger texts were developed for eight stages of conversation (those in Table 2 plus a greeting and farewell exchange). The ghost protagonist (to be played by the chatbot) was imagined as a regretful ghost, reflecting on the events from the play (see Fig. 1, which shows answers for key event 1 'the protagonist was a king').

Once both Segway and Cliffhanger texts had been written, they were formatted and uploaded to the Storybox system for access by the live training environment to develop couplets for the conversations.

In Phase 2, 'The Ghost' chatbot training took place over the first two days of the Playful Learning 2018 academic conference, with a qualitative evaluation taking place on the third day. The conference was attended by educators and trainers from schools, colleges and professional environments. During breaks and lunchtimes, attendees were invited to try out The Ghost using a PC set up in the coffee area, and a

second PC for the trainer was hidden close by. Participants were given information about the system and asked to hold a conversation playing the role of a student wishing to find out the story. 25 separate conversations were recorded with the trainer playing the Ghost character. On the final day of the conference, the system was switched to chatbot mode and attendees were invited to converse with The Ghost storytelling chatbot. Following the conversation with the system in chatbot mode, participants were invited to complete a questionnaire to report their views on the usability of the system.

7. Results and Discussion

As reported above, the live training phase of the development resulted in 25 separate training conversations conducted with attendees at the conference. Following the two training days, 15 adult participants conducted conversations with the live chatbot called The Ghost. The participants were educators, with expertise in the use of technology in educational settings, who were asked to play the role of students based on their knowledge of learner behaviour. Participants had no prior experience of The Ghost chatbot. Anonymous online questionnaires were completed by 9 of the 15 participants who evaluated The Ghost chatbot. The questionnaires were designed to assess the extent to which the participants believed The Ghost to be a convincing story character within a learning context.

All respondents described the Ghost character as either convincing (4) or partly convincing (5). When asked to describe in more detail the aspects of the experience that were convincing, participants commented on the accuracy of some of the responses from the chatbot:

The answers made sense in the conversation

Some parts of the conversation flowed well enough that it seemed natural

Several answers were spookily accurate and meaningful.

Sometimes the answers made a lot of sense and seemed to be well related to my question or comment

Asked to comment on what they found unconvincing, the respondents listed points at which the chatbot did not answer them and times when it repeated itself. (From researcher observation of the session, participant comments about repetition seemed to refer to an isolated out-of-sequence text caused by an error in an earlier training session that came up more than once in answers to users):

Occasionally the ghost did not respond and I was unsure what to do next

When you got stuck in a loop and you got the same response several times in a row

When the answers repeated themselves and were not quite on the mark.

When asked to reflect on the experience of conversing with The Ghost, participants all agreed that their conversation with The Ghost had been *interesting*. It was also variously described as *surreal*, *disconcerting* and *uncanny and odd* at the points at which it seemed to really understand them and respond correctly.

Overall, the results show that participants found conversing with The Ghost character to be a convincing experience that could engage students. Given the small amount of data that the system used for training (under 200 couplets and 25 conversations in total), the level of clarity that the chatbot was able to maintain in providing answers to participants was good. This suggests that using the Storybox framework, it is possible for chatbots to be deployed more easily in a number of expert applications with less investment of time and effort in order to produce an effective agent. As other chatbot studies cited in Section 1 illustrate, systems with 100,000+ sets of conversation data using advanced machine learning methods struggle to maintain sense and realistic responsiveness even during short conversations with users.

However, based on participant feedback there are straightforward enhancements that can be made to improve the system:

1. Install an automatic response or set of responses to answer the user at times when the database does not find a suitable match for a question. This might take the form of a prompt to phrase their question differently or some other indication that it does not understand.

2. Add a data cleansing and refinement stage to the Storybox chatbot methodology following the live training sessions in order to delete erroneous responses and to improve or adapt incomplete or inadequate responses written in the live environment.
3. In order for a non-technical user to write the story script for training, a future iteration of the Storybox system could provide automation of the more technical elements of development. This could allow users themselves to create SCAs. This would allow them to develop new and increasingly topical English Language skills by taking a creative approach to producing a novel interactive document. It could also provide opportunities to teach students AI and be part of a wider pedagogical trend (e.g. Ma, 2019) that seeks to demystify algorithmic culture.

8. Conclusion

This paper has described Storybox, a novel methodology and architecture for developing a storytelling chatbot for possible use in a broad range of expert settings. Storybox has been designed to support the faster development of chatbots for storytelling, aiming to overcome the barrier of extensive development time by incorporating a new way of structuring knowledge, and a live training phase which populates general conversational questions and answers from real interactions with human experts. The structured real time interaction between the subject expert and training participants used in the Storybox approach overcomes the problem noted in the literature regarding the domain expert also needing to possess technical CA system knowledge. Also proposed was D-PAF, a new framework for structuring chatbot responses suited to the storytelling context, which overcomes the problem of digressing away from the story by producing an automatic three segment chatbot response which responds to the user input but also progresses the story being told. An initial pilot study was described, which followed to Storybox methodology to develop a chatbot called The Ghost that was trained and evaluated by expert educators. The study showed that the Storybox methodology and D-PAF do provide an effective toolkit for creating storytelling chatbots with a relatively small group of testers and trainers. This overcomes the challenges of traditional chatbot systems that are either not designed to tell stories or that require large amounts of preparation before they can be deployed in live environments.

The key benefit of this toolkit to educators in the classroom and external learning settings, such as museums and cultural venues, is to make storytelling chatbots a feasible teaching and learning tool for users who are usually reluctant to read and write using other methods. Storybox can be used for rapid chatbot development and deployment in a live learning environment with considerably less data resources than has previously been possible. The D-PAF framework successfully enforces the linear prerogative of the storyteller to tell one story whilst also listening and responding to the questions asked by the user. The evaluation documented in this paper suggests that the blended approach taken by Storybox, combining live conversational training with branching tree narrative and linear narrative methods is an effective new method for structuring the data required for telling stories.

In future, a study will be conducted to investigate how this lightweight approach could be used in an educational setting to provide opportunities for reluctant readers and writers to interact in a novel way with educational source texts such as Hamlet, the text used in the evaluation. Part of this investigation could involve the co-design of a system for authoring the sections of the chatbot that require pre-population, so that no technical knowledge is required to make and train a new chatbot. This would allow teachers and learners to focus on the creative writing tasks involved in the process. The study should also investigate the scope for children and young people to take part in the pre-population and training stages of creating a chatbot. This could deepen the learning benefits of the system and help students to better understand the way in which conversational digital devices artificial intelligence systems are designed, developed and trained.

Acknowledgements

The Storybox project was developed as part of an AHRC-funded Creative Economy Engagement Fellowship awarded by the North West Consortium Doctoral Training Partnership.

References

- Ahn, J., Watson, P., Chang, M., Sundararajan, S., Ma, T., Mukhi, N. and Prabhu, S. (2017). ‘Wizard’s Apprentice: Cognitive Suggestion Support for Wizard-of-Oz Question Answering.’ In André, E., Baker, R., Hu, X., Rodrigo, Ma. M. T., and du Boulay, B. (eds) *Artificial Intelligence in Education*. Springer International Publishing (Lecture Notes in Computer Science), pp.630-635.
- Arsovski, S., Osipyan, H., Oladele, M.I. and Cheok, A.D. (2019). Automatic knowledge extraction of any Chatbot from conversation. *Expert Systems with Applications*, 137, pp.343-348.
- BOP Consulting (2015). *Literature Review: The impact of reading for pleasure and empowerment*. The Reading Agency. [Online] [Accessed on 10th July 2019] <https://readingagency.org.uk/news/The%20Impact%20of%20Reading%20for%20Pleasure%20and%20Empowerment.pdf>.
- Brewer, W. F. and Lichtenstein, E. H. (1982). ‘Stories are to entertain: A structural-affect theory of stories.’ *Journal of Pragmatics*, 6(5-6) pp. 473-486.
- Campbell, J. (1949). *The hero with a thousand faces*. MJF Books, New York.
- Crawford, C. (2004). *Chris Crawford on Interactive Storytelling*. New Riders Games.
- Clark, C. (2018). Writing enjoyment, behaviours and attitudes in 8 to 11-year-olds in 2017/18: Findings from our Annual Literacy Survey.
- Coronado, M., Iglesias, C.A., Carrera, Á. and Mardomingo, A. (2018). A cognitive assistant for learning java featuring social dialogue. *International Journal of Human-Computer Studies*, 117, pp.55-67.
- Crockett, K., James, O.S. and Bandar, Z. (2011). June. Goal orientated conversational agents: applications to benefit society. In *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications* (pp. 16-25). Springer, Berlin, Heidelberg.
- Crown (n.d.). *National curriculum in England: English programmes of study* GOV.UK. [Online] [Accessed on 10th July 2019] <https://www.gov.uk/government/publications/national-curriculum-in-england-english-programmes-of-study/national-curriculum-in-england-english-programmes-of-study>.
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22(5), 811-817. doi:10.1017/S1351324916000243.
- De Kleijn, R., Wijnen, M. and Poletiek, F. (2019). The effect of context-dependent information and sentence constructions on perceived humanness of an agent in a Turing test. *Knowledge-Based Systems*, 163, pp.794-799.
- Filipczyk, B., Gołuchowski, J., Paliszkievicz, J. & Janas, A. (2016). Chapter 12 - Success and failure in improvement of knowledge delivery to customers using chatbot—result of a case study in a Polish SME. In J. Liebowitz (Ed.), *Successes and Failures of Knowledge Management*. Morgan Kaufmann, 175-189.
- Gadney, G. (2018) *An Era of Character, Context and Conversation - Guy Gadney*. Medium. [Online] [Accessed on 10th July 2019] <https://medium.com/@GuyG21/an-era-of-character-context-and-conversation-88a641eabdd2>.
- Gerrig R., Love J., McKoon G. (2009). Waiting for Brandon: How readers respond to small mysteries. *Journal of Memory and Language*. 60:144–153.
- Graesser, A.C., Cai, Z., Morgan, B. and Wang, L. (2017). Assessment with computer agents that engage in conversational dialogues and triologues with learners. *Computers in Human Behavior*, 76, 607-616.
- Graesser, A. C., Li, H., & Forsyth, C. (2014). Learning by Communicating in Natural Language With Conversational Agents. *Current Directions in Psychological Science*, 23(5), 374–380. <https://doi.org/10.1177/0963721414540680>.
- Herbert, D. and Kang, B.H. (2018). Intelligent conversation system using multiple classification ripple down rules and conversational context. *Expert Systems with Applications*, 112, pp.342-352.
- IBM (2019). *AI for Customer Service | IBM Watson*. Available: <http://www.ibm.com/watson/ai-customer-service> [4 Mar. 2019].

- Jones, A. and Issroff, K. (2005). Learning technologies: Affective and social issues in computer-supported collaborative learning. *Computers & Education*, 44(4), 395-408.
- KPMG and National Literacy Trust (2013). Youth Literacy and Employability Commission: The report of the All-Party Parliamentary Literacy Group. [Online] [Accessed on 10th July 2019] https://cdn.literacytrust.org.uk/media/documents/2013_01_01_free_other_-_Youth_Literacy_and_Employability_Commission_final_report.pdf.
- Kumar, R. and Rosé, C. P. (2014) Triggering effective social support for online groups. *ACM Trans. Interact. Intell. Syst.* 3, 4, Article 24 (January 2014), 32 pages.
- Latham, A. M., Crockett, K. A., McLean, D. A., & Edmonds, B. (2012). A conversational intelligent tutoring system to automatically predict learning styles. *Computers & Education*, 59(1), 95–109.
- Latham, A., Crockett, K. and McLean, D. (2014). An adaptation algorithm for an intelligent natural language tutoring system. *Computers & Education*, 71, pp.97-110.
- Leuski, A. and Traum, D. (2011) 'NPCEditor: Creating Virtual Human Dialogue Using Information Retrieval Techniques.' *AI Magazine*, 32, June, pp. 42–56.
- Lewin, C. (2000). Exploring the effects of talking book software in UK primary classrooms. *Journal of Research in Reading*, 23(2), 149-157. Retrieved from <https://search-proquest-com.mmu.idm.oclc.org/docview/85520317?accountid=12507>.
- LogMeIn Inc. (2019). *Chatbots for B2B Lead Generation: How to Deliver Personalized Conversations at Scale | Bold360*. Available: <https://www.bold360.com/resources/articles/whitepapers/chatbots-for-b2b-lead-generation-how-to-deliver-personalized-conversations-at-scale> [4 Mar. 2019].
- Love, J., McKoon, G. & Gerrig, R. (2010). Searching for Judy: How Small Mysteries Affect Narrative Processes and Memory. *Journal of experimental psychology. Learning, memory, and cognition*. 36. 790-6. 10.1037/a0018989.
- Loyall & Bates (1997) .Personality-Rich Believable Agents That Use Language. Proceedings of the First International Conference on Autonomous Agents. Marina del Rey, California.
- Ma, M., Coward, S., & Walker, C. (2017). Question-answering virtual humans based on pre-recorded testimonies for holocaust education. *Serious Games and Edutainment Applications: Volume II*.
- Mateas, M. and Stern, A. (2003). *Façade: An Experiment in Building a Fully-Realized Interactive Drama*. Game Developers Conference, San Jose.
- Mateas, M. and Stern, A. (2007). *Façade*. Electronic Literature Collection Volume Two. Available: <http://collection.eliterature.org/2/index.html> [15 Jul. 2019].
- McGrane, J., Stiff, J., Baird, J.-A., Lenkeit, J. and Hopfenbeck, T. (2017) *Progress in International Reading Literacy Study (PIRLS): National Report for England*.
- McKee, R. (1997). *Story: Substance, Structure, Style and the Principles of Screenwriting*. London: Methuen Publishing Ltd.
- McNeal, M.L. & Newyear, D. (2013). "Chapter 1: Introducing chatbots in libraries", *Library Technology Reports*, vol. 49, no. 8, pp. 5.
- Miller, C. H. (2004). *Digital Storytelling: A Creator's Guide to Interactive Entertainment*. Taylor & Francis.
- Montenegro, J.L.Z., da Costa, C.A. and da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Systems with Applications*, 129, pp.56-67.
- Murray, J. H. (1997). *Hamlet on the holodeck: the future of narrative in cyberspace*. MIT Press.
- Neumann, A.T., de Lange, P. and Klamma, R. (2019). December. Collaborative Creation and Training of Social Bots in Learning Communities. In 2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC) (pp. 11-19). IEEE.
- O'Shea, J., Bandar, Z., & Crockett, K. (2011). Systems engineering and conversational agents. In A. Tolk, & L. C. Jain (Eds.), *Intelligence-based systems engineering*. Berlin Heidelberg: Intelligent Systems Reference Library 10, Springer-Verlag.
- Propp, V. I. (1968). *Morphology of the Folktale: Second Edition*. University of Texas Press.
- Ryan, M. 2006, *Avatars of story*, University of Minnesota Press, Minneapolis.

- Sarıca, H.Ç. and Usluel, Y.K. (2016). The effect of digital storytelling on visual memory and writing skills. *Computers & Education*, 94, pp.298-309.
- Sebastian, J. and Richards, D. (2017). Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior*, 73, pp.479-488.
- Shiro, M. S. (2004). *Oral storytelling & teaching mathematics: pedagogical and multicultural perspectives*. London: SAGE, 2004
- Short, E., 2000. Galatea. The Interactive Fiction Database. [online] [Accessed on 10th July 2019] <https://ifdb.tads.org/viewgame?id=urxrv27t7qtu52lb>.
- Sontag, S. (2005). At The Same Time ... (The Novelist and Moral Reasoning), *English Studies in Africa*, 48:1, 5-17.
- Tegos, S., Demetriadis, S. and Karakostas, A. (2015). 'Promoting academically productive talk with conversational agent interventions in collaborative learning settings' *Computers & Education*, 87, September, pp. 309–325.
- Tegos, S., & Demetriadis, S. (2017). Conversational agents improve peer learning through building on prior knowledge. *Educational Technology & Society*, 20(1), 99+. Retrieved from <https://link-gale-com.mmu.idm.oclc.org/apps/doc/A482056355/AONE?u=mmucal5&sid=AONE&xid=a584e477>.
- Wang, S.L. and Wu, P.Y. (2008). The role of feedback and self-efficacy on web-based learning: The social cognitive perspective. *Computers & Education*, 51(4), 1589-1598.
- Wang, Z., Wang, Z., Long, Y., Wang, J., Xu, Z. and Wang, B. (2019). Enhancing generative conversational service agents with dialog history and external knowledge. *Computer Speech & Language*, 54, pp.71-85.
- Windrip-Fruin, N. (2009). *Expressive processing: digital fictions, computer games, and software studies*. London: The MIT Press.
- Winston, J., & Royal Shakespeare Company. (2015). *Transforming the teaching of shakespeare with the royal shakespeare company*. London: Bloomsbury Arden Shakespeare.
- Yan, Z., Duan, N., Bao, J., Chen, P., Zhou, M. and Li, Z. (2018). Response selection from unstructured documents for human-computer conversation systems. *Knowledge-Based Systems*, 142, pp.149-159.
- Yang, Y.T.C. and Wu, W.C.I. (2012). Digital storytelling for enhancing student academic achievement, critical thinking, and learning motivation: A year-long experimental study. *Computers & Education*, 59(2), pp.339-352.
- Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D. and Weston, J. (2018). 'Personalizing Dialogue Agents: I have a dog, do you have pets too?' *arXiv:1801.07243 [cs]*, January.

Highlights

- Storybox methodology offers new model for rapid development of storytelling chatbots
- New framework (D-PAF) provides structure for storytelling dialog
- Small knowledge base led to convincing chatbot conversations during evaluation
- Rapid training method captured conversation from live sessions