The University Of Sheffield.

Scene Annotation Guidelines

Authors:

Tarfah Alrashid, Prof. Robert Gaizauskas

Supervisor:

Prof. Robert Gaizauskas

June 2021

Table of Contents

1 Introduction

To date, there has been a lot of research interest in integrating vision and language for different tasks. Image captioning for instance, is the task of generating natural language descriptions for a given image. Models are trained using datasets having texts aligned with images that are manually annotated. Another example of an image-based task is automatic text illustration, which is the task of automatically finding a suitable image for a given segment of text.

Such research integrating vision and language generally tries to solve image-based tasks, by relying on images aligned with texts. The problem is twofold: (1) many captions associated with images do not just describe the image; (2) the volume of captions overall is not great.

Therefore, we propose exploring how text not associated with images can be used to help in this type of research. We can find such text in narratives and drama, because we get descriptions of setting in fiction. We may get visual descriptions of things in other writing also, for example biography. So we propose to investigate ways of utilising these other purely linguistic image-based resources to help in tasks relating to images.

Our research aims to explore the idea that scenes are divisions found in all narratives which, amongst other things, have to do with physical setting. Usually there are descrip- tions that enable the reader to note scenes have changed. Automatic scene segmentation could be useful for purposes of illustration and alignment. This project proposes the idea of automatic segmentation of narrative text (stories) into scenes. In order to carry out this task, we first provide a definition of what a scene is. Then, we define guidelines for scene annotation that help in identifying scene change and scene boundaries. Automatic segmen- tation of text into scenes could help in building a text corpus of scenes that could be used by systems converting scenes into images. In addition, scenes could be clustered and common visual elements identified for scenes clustered together. If scene type labels can be given to these clusters, then these labels could be attached to images whose visual elements model these clusters.

1.1 Overview of the Report

The report is structured as follows:

- Section 2 gives a brief introduction to SceneML.
- Section 3 explains the annotation process.
- Section 4 explains how to use the annotation tool.

2 SceneML

2.1 The Annotation Framework

Taking into consideration the literature related to scene definition, we treat a scene as a unit of a story in which the elements: time, location, and main characters are constant. Any change in these elements indicates a change of scene. A scene is an *abstract* discourse element. It consists of a location or setting, a time and characters who are involved in the events that take place in the scene. These elements that constitute a scene exist in the real or fictive world (*storyworld,* as per narrative theory) that the narrative revolves around. “The scene itself is an abstraction away from the potentially infinitely complex detail of that real or fictive world, a cognitive construct that makes finite, focussed narrative possible" [2].

A scene in a textual narrative may consist of one or more *scene description segments* (SDS). An SDS is “a contiguous span of text that, possibly together with other SDSs, ex- presses a scene" [2]. Typically a scene consists of one SDS, unless the scene introduces other scenes such as scenes about past (memories) or future events, in which case the textual realisation of one scene may be split with text from another scene embedded within it, resulting in non-contiguous SDSs for the embedding scene. Furthermore, SDSs can be in the form of “spatially distinct locations that are topologically contained within or connected to the embedding SDS, or if the author is employing the narrative device of rotating between multiple concurrent scenes each of which is advancing a distinct storyline (a common technique in action movies)"[2].

It is important to define what each element of a scene (*characters*, *time*, *location*) is to facilitate the annotation process and to make it easier to detect scene changes if any of the elements changed. Scene changes have been studied by researchers such as [1] and others. Studying which of the scene elements (*characters*, *time*, *location*) changes in a scene can be considered a contribution to previous studies. Here we propose to adopt the definitions, and annotation standards, for the elements (*time*, *location*, and *spatial entities*) from Iso-TimeML and Iso-Space. As for characters, we will adopt the definition and annotation standards for named entities of type person from the ACE program which has been recently used in in the TAC 2018 entity discovery and linking task [2].

The previous standards will facilitate the annotation process for all mentions of times, locations/spatial entities and persons represented in the text. Nonetheless, we are inter- ested in the relations between the specific entities that construct a specific scene (i.e., the *characters*, *time*, *location* that construct a specific scene) [2].

2.2 SceneML Elements

SceneML elements are categorised into two main categories:

1. **Entities:** entities consists of scenes, SDSs, character, time, and locations
2. **Relations:** “scene-scene narrative progression links".

Scenes

Scenes are the main element in SceneML. A scene has attributes: id, time and location, these attributes are unique for each scene. They also include a list of character sub- elements as there may be more than one character for each scene [2].

SDSs

Scene description segments (SDSs) are the actual strings of text that compose the scene. SDSs cannot belong to more than a scene, but a scene can be composed of multiple SDSs. They include the following attributes: id and scene_id that is the id of the scene that the SDS belongs to, these attributes are unique to each SDS [2].

Time

Time elements used here are the ones developed by ISO-TimeML, they include an id attribute and a text segment. Time could also include the time of the storyworld and it is represented by the attribute base [2].

Location

Like time, we use the location element developed in ISO-Space; this also includes an id attribute that is unique for each location and a text span [2].

Character

Here we use the named entity or type person from the ACE English Annotation Guidelines for Entities, the only difference here is that animals and non-humans can be considered as characters. They have an id attribute that is uniquely assigned for each character and a text segment. Character might include the type attribute of the ACE specification [2].

Narrative Progression Links

Narrative progression links (nplinks) link between any two textually adjacent scenes, these links have different types depending on the type of temporal transition between the two scenes. So far we have four types of links [2]:

- sequence links are assigned when the scene change happen because of a change in location e.g. a character moves to another place
- analepsis when there is a flashback in the scene e.g. memory of the past
- prolepsis (or flashforward) when then we are taken forward in time
- concurrent links are assigned between two scenes when the transition happen because there is another thread of the story happening at the same time so the transition take us to different charters and different place but the same time

3 The Annotation Process

Novels and stories usually consist of a set of scenes. A scene as defined in literature is a unit of a story in which the elements: time, location, and the main characters are constant. Any change in these elements indicates a change of scene. A scene is an abstract discourse element, i.e. it is not explicitly indicated in the text. It consists of a location or setting, a time and characters who are involved in the events that take place in the scene. A scene in a story may be realised by one or more scene description segments (SDS). An SDS is "a contiguous span of text that, possibly together with other SDSs, expresses a scene"[1]. Typically a scene consists of one SDS, unless that scene involves other SDSs of other scenes such as scenes about past (memories) or future events. Furthermore, SDSs can be in the form of "spatially distinct locations that are topologically contained within or connected to the embedding SDS, or if the author is employing the narrative device of rotating between multiple concurrent scenes each of which is advancing a distinct storyline (a common technique in action movies)"[1].

Example: Figure 1 shows an illustration of how scenes should be connected through relations. *Scene1* consists of two non-adjacent SDSs, *SDS1* and *SDS3*. While *Scene2* consists of only one SDS, *SDS2*. *Scene1* and *Scene2* are connected by a flashback relation.

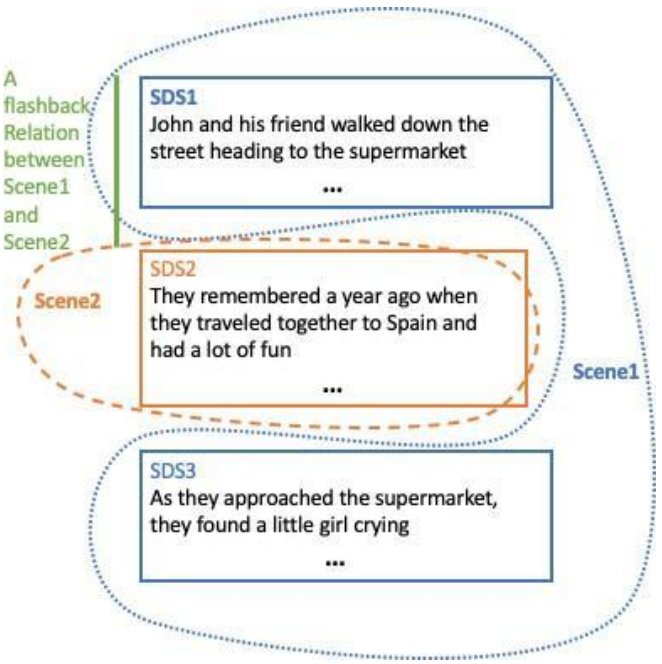

Figure 1: Two scenes, *Scene1* and *Scene2* connected with a flashback relation.

Textually adjacent scenes, i.e. scenes with textually adjacent SDS's, should be linked together through progression links. Note that these links are links between scenes not SDSs. We identify four types of progression:

- sequence, when one scene follows on from another, e.g. when characters move from

one location to another;

- analepsis (or flashback), when we are taken to another, earlier time and possibly other details such as location and characters change as well;
- prolepsis (or flashforward), when we are taken forward in time;
- concurrent, when we are taken to another location with different characters, where another thread of the story is developing at the same time as the textually preceding scene.

The annotation tool does not support the creation of abstract entities such as scenes. Therefore to capture the information that multiple SDSs belong to the same scene, we will annotate a relation between SDSs that are in the same scene ("is_in_the_same_scene_as"). Furthermore to capture progression relations between scenes, we will annotate these relations between the initial SDSs of related scenes, with the understanding that really the relation holds between scenes and not SDSs. Figure 2 shows how to annotate the example in Figure 1 with the BRAT annotation tool. The flashback relation is connecting *SDS1* from *Scene1* with *SDS2* from *Scene2*. *SDS1* is connected to *SDS3* with a "is_in_the_same_scene_as" relation to indicate that they are in the same scene. It is im- portant to define what each element of a scene (characters, time, location) is to facilitate the annotation process and to make it easier to detect scene changes if any of the elements changed.

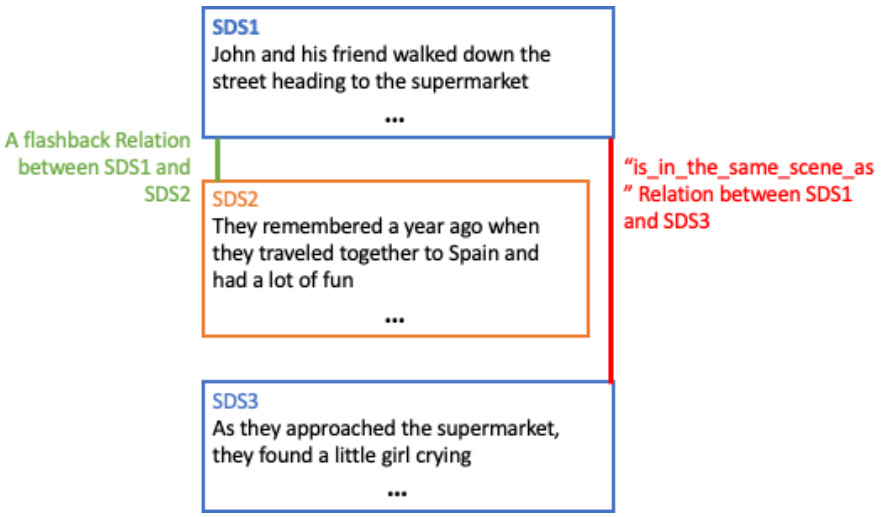

Figure 2: How scenes and SDSs are annotated with the annotation tool.

You are required to annotate SDSs in the text provided, then link these SDSs together if you see that multiple SDSs compose a scene. Note that each SDS must belong to only one scene, but every scene can contain one or more SDSs. In addition, you are required to annotate the elements: time, location, and characters. Please note that the change of these elements indicates a change in the whole scene not a change in the SDSs. Please also annotate the relations between scenes as explained above.

3.1 Detailed Annotation Guidelines

Below you will find more detailed explanations of each of these elements and instructions for the annotation process.

Entities to be Annotated:

Scenes

Scenes are the main element in SceneML. A scene has elements time and location, which are unique for each scene. A scene also includes a list of character sub-elements as there may be more than one character for each scene. Please note that this is an abstract element, which means that this element is not annotated explicitly in the text.

Then how can one annotate a scene if it is an abstract element? you can annotate a scene by: (1) if the scene has only one segment (SDS) then all you have to do is to annotate the segment as (SDS), (2) if the scene has multiple segments (SDSs), then annotate each one of these segments as (SDS) separately and then link them together with *is_in_the_same_scene_as* relation. By linking these segments together we indicate that they all belong to the same scene.

So all you have to do here is to build a conceptual understanding of where the scenes are in the text and annotate the SDSs that constitute these scenes. While you do that, you also need to pay attention to the three main elements that compose the scene, and once you notice a change in any of these elements you need to end annotating the current SDS that compose the current scene and start a new one by following points one and two above.

SDSs

Scene description segments (SDSs) are the actual strings of text that compose the scene. An identified SDS cannot belong to more than one scene, but a scene can be composed of multiple SDSs. In order to identify an SDS, you need to make sure that the 3 elements of a scene (time, location, characters) persist across that continuous span of text. The moment you notice a change in any of the elements, you need to consider a new scene and the end of the current SDS. Sometimes, the time entity might not be stated literally in the text; it can be inferred from the tense used, or it might be only mentioned at the beginning of the story. It is OK for an SDS to not have a time entity explicitly mentioned as long as the time persists from a previous scene; similarly for location. Please note that each SDS should be at least one sentence long.

Scene Boundaries and scene shifts:

Here you will find more information on how to end the current scene (SDS), and begin a new one. And what could be considered a sufficient change for a scene shift.

- **If there is a location change that results in a scene shift, where exactly should we stop the current SDS, and start the following one?**
  Transitioning sentences stating that a person is arriving at a new location or leaving the current location should be handled one of two ways: (1) if the location is preceded by a verb like *arrived*, *got to,* … etc. then we can consider the sentence to belong to the following, new SDS (2) if the location is preceded by a verb like *left*, *got out of,* …, etc.) then we can consider the sentence to belong to the current SDS.
- **Changing locations: when to consider a change in a location, if you are unsure about changing the scene?** This means if the location is changed but you are not sure if that change in the location (e.g. very minor change) will lead to a scene change.
  - If there is a very minor location change that you feel does not lead to a change in the scene, e.g. if the characters of the current scene only move from one side of the room to the other without exiting the current room, then this minor location change should be ignored, and there is not a shift in scenes.
  - Try to ask yourself, if you were to design sets for a play version, would you need a different backdrop for it?.
- **Changing characters:** when reading the whole text first, you'll be able to build a general idea about who are the characters that contribute more to the story and who are the characters that do not have a major impact on the scenes of the story. When major characters enter the current scene or leave the scene, you should end the scene at the current SDS and start a new one. Nevertheless, if minor characters that do not have a major role in the scene, have any change in their situation (entering, or leaving), it should not affect the current SDS of the scene, and thus the scene should continue.

Non-scene segments: please note that not all segments (i.e. paragraphs/ sentences/ phrases ... etc) need to be annotated as belonging to an entity. Some segments can be left without annotation if you see that they do not belong to a scene or do not create a scene. Some of these non-scene segments may reflect the author's general view of something, or may be a description of a character or a place or some period of time prior to setting that the rest of the story takes place in. So far we have identified three types of a non-scene segment; there may be more:

1. general philosophising or opinion segments -- these do not mention specific characters or events but rather offer the author's view on some topic or another;
2. background information segments -- they may lack precise details of time or location, are clearly set before the current narrative begins, and provide insight into characters or past events of relevance for understanding the story that is about to begin;
3. narrative summary or narrative catchup -- where various events, usually non-pivotal, are summed up in a relatively compressed passage that lets the reader know how events have moved on in between two scenes that are presented in greater detail, without precisely indicating all the details of character, location and time. Events themselves in these passages may not be distinguished from each

other, if the intention is to indicate a trend or only the outcome.

For example, the first two paragraphs of chapter 1 of the novel Adventures of Sherlock Holmes are non-scene segments because they contain background information. They describe the characteristics of the character *Sherlock Holmes* and the relationship between him and his friend *Watson*. There is nothing actually happening at these segments, so they should not be annotated as scenes. You can find the novel here: https://www.gutenberg.org/files/1661/1661-h/1661-h.htm#chap01. In addition, the first two paragraphs of the chapter 1 of the novel *A Tale of Two Cities* are non-scene segments (background information) as they describe the era and the character of that era, but no specific events are related. You can find the novel here: https://www.gutenberg.org/files/98/98-h/98-h.htm. In another example, the first two paragraphs of chapter 1 of Pride and Prejudice - *It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife… etc.* And the first two paragraphs of chapter 1 of *A Tale of Two Cities* are non-scene general philosophising segments. You can find *Pride and Prejudice* here: https://www.gutenberg.org/files/1342/1342-h/1342-h.htm#link2HCH0001. The first paragraph of chapter 3 of *Pride and Prejudice* is a nice example of a narrative summary non-scene segment.

As you can see from the examples above most of the non-scene segments occurred at the beginning of a story or the beginning of a chapter, this is what we have encountered so far but non-scene segments can be embedded in text. If you find a non-scene segment embedded in text that is less than a full sentence, ignore the segments and consider it to be part of the current SDS segment.

Scene Transitioning Segments (ST)

These are text segments that indicate a character or characters are moving from one location to another, where this transition is neither clearly related to the previous scene location nor the following scene. In addition, there is not any event of narrative significance happening during the transitioning segment other than the transition. These segments should not contribute to the narrative of the story (e.g. if there is a conversation happening between characters of the story, they should not speak important information related to the story) . For example,

> *And so in ten minutes I had left my armchair and cheery sitting-room behind me, and was speeding eastward in a hansom on a strange errand, as it seemed to me at the time, though the future only could show how strange it was to be.*

The above segment (*Adventures of Sherlock Holmes* chapter 6) is annotated as a transitioning segment (ST), it shows that the character Watson has left the previous scene location (his house) and is en route to the next scene location.

Time

Here you need to annotate the text span that refers to the time of the scene (i.e. the time the events of the scene are happening). Below you will find more examples of what could be considered a time of a scene and more instructions on the annotation process.

Time spans could be:

1. Noun (including Proper Nouns): e.g. today, Thursday

2. Noun Phrase (NP): e.g. the morning, Friday night, the last two years
3. Adjective: e.g. current
4. Adverb: e.g. recently
5. Adjective or Adverb Phrase: e.g. half an hour long, two weeks ago, nearly a half- hour

Please be aware to annotate the time entity that the scene events are happening in. Some scenes might have multiple mentions of time entities (they could be mentioned in a question or in a description of something else) although only one will be considered as the time of the scene. For example, consider this question that was asked by a character in a children's story: “Do you not have good fabrics in the future?” I asked, eyeing a box that seemed to be stuck to the wall" here the word future should not be annotated as the time of the scene in as it does not refer to the time the scene events are happening.

Examples of a time entity:

- **ONE TIMEX tag will be used when there is no intervening token between temporal terms:**
  - twelve o’clock midnight
  - Friday evening
  - 8:00 p.m. Friday
  - Tuesday the 18th
  - November 1943
  - Fall 1998
  - this year’s summer
- **ONE TIMEX tag will be used when certain prepositions appear within temporal expressions:**
  - the second of December
  - summer of 1965
  - ten of two ten minutes to three
  - half past noon
  - eleven in the morning
- **TWO TIMEX tags will be used when sequences of two temporal expressions that are ordered one relative to the other. They generally involve the use of temporal prepositions and conjunctions like from, before, after, following, prior to, etc. :**
  - I’m leaving on vacation [two weeks] from [next Tuesday]. John left [2 days] before [yesterday]
- **TWO TIMEX tags will be used when sequences of two temporal expressions that can be related by a temporal link:**
  - I tutored an English student [some Thursdays] in [1998]. The concert is at [8:00 p.m.] on [Friday]. The concert is [Friday] at [8:00 p.m.]

Location

Like time, you need to annotate the text span that refers to the location of the scene (i.e. the location the events of the scene are taking place in). Below you will find some examples of

locations, and some annotation cases you need to be aware of.

Some examples of possible locations of scenes:

- Building
- United States
- West Tikrit

Some notes in annotating locations:

- Please make sure when annotating a location that the location is the location of the current scene, as some scenes may have mentions of more than one location, though only one of them is the location where the scene events happen. For example, " "You just wait here until we're ready to take you home," he said, as he floated towards one of the exit tunnels". If the word "home" were to be tagged as the location of the scene that would not be correct as events of the scene do not take place at "home" — it is only mentioned in the conversation.
- Phrases like "centre of the room" was tagged as being the location of the scene, it is not completely wrong but "room" is supposed to be tagged alone as the (Location) of the scene instead of tagging the words (centre, of, and the) as it does not matter where in the room the scene is happening as long as it is in the room the scene would not change.
- Similarly, all the words in the sub sentence "back into the corridor" were annotated as a location and that is not correct, only the corridor should have been tagged as a location.

Character

Text spans that refer to characters need to be annotated, characters could be humans, non humans (e.g. animals), and subjects. Below some specific cases in annotating characters:

- Only the first mention of each character should be annotated in each scene. Characters, could be mentioned several times in every scene or sds, but only the first mention should be annotated.
- Sometimes subjects (e.g. he, she) could be used to a character that is already annotated within the same scene, then this subject need not to be annotated. For example: John is playing football with his friend. He is a good football player. If the two sentences belong to one scene, then we only annotate John. But if the two sentences belong to two different scenes, then we annotated John as an actor of the first scene and He as an actor of the second scene.
- If the subject of a certain character is mentioned before its character in the same scene, then we do not annotate the subject, we annotate the name of the character instead. The subject is mentioned before its name, when the name is mentioned in a previous scene, so in this case we annotated the first mention of the name in the scene. If the name is not mentioned in the scene at all, then we annotate its first mention of its subject.
- When multiple characters appear next to each other (e.g. Biff, Boff and Skip) each one of them should be tagged alone as a character, do not annotate all of them as one segment. Because then the word and will be also tagged as a character and that is not

true.

3.2 How to annotate

Start by reading the whole text first without annotating to gain a whole understanding of the story and events. Then, read it again and annotate all SDSs and within each SDS annotate the three entities if they exist as explained in the previous section.
If no time is given in the current scene, add a note to the scene with the time being as the previous scene.

4 BRAT (Annotation tool)

- Go to the page: http://sceneml.shef.ac.uk/brat/#/your name/

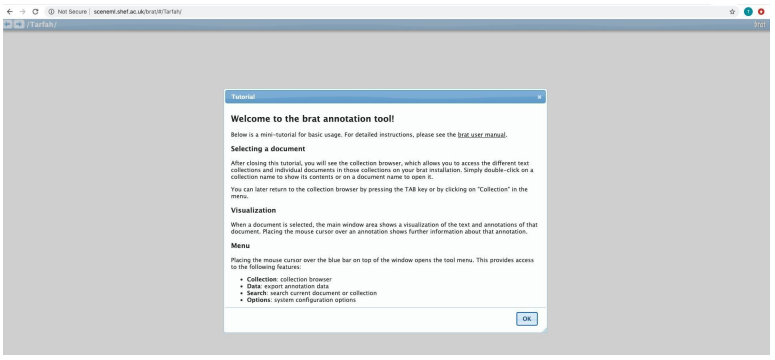

Figure 3

- A popup window will appear that contain all the files in you page, select the file you want to annotate and click OK.

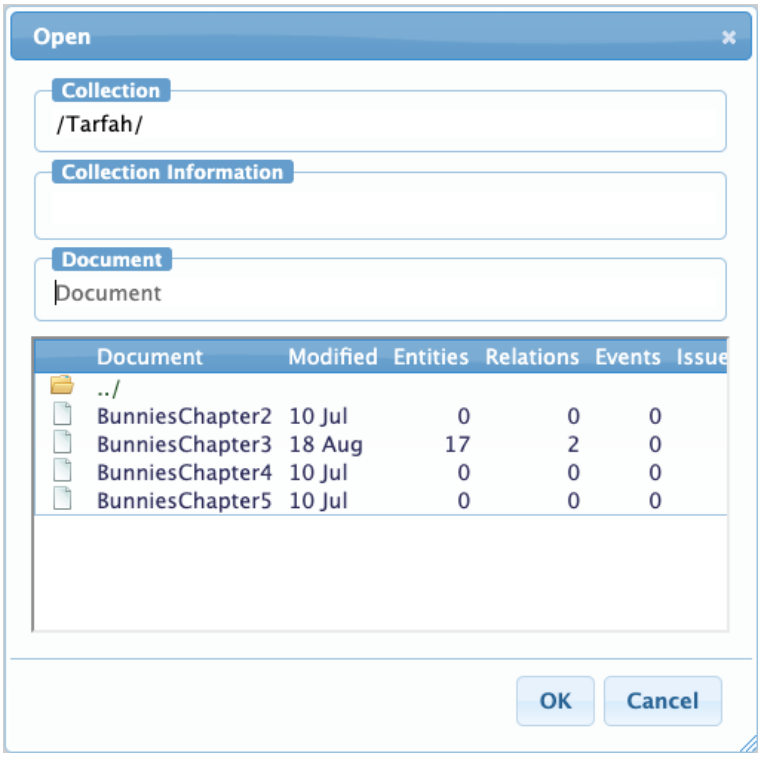

Figure 4

- In order to annotate, you'll need to login first. Username and password has been sent to your email. Hover the mouse on the login icon in the top bar and log in (see Figure 5).

Figure 5

- Select the text span you want to annotate as in Figure 6

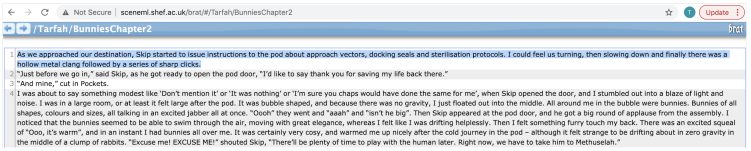

Figure 6

- A pop up menu will appear with the all entity types, select the entity type you want and click OK (see Figure 7)
- To connect two SDSs together;


New Annotation
Text
As we approached our destination, Skip started to issue instructions to the pod about approach vectors, docking seals and sterilisation protocols. I could feel us turning, then slowing down and finally there was a hollow metal clang followed by a series of sharp clicks.
Search
Google, Wikipedia
Entity type
SDS
Scene
Character
Location
Time
Notes
OK
Cancel

Figure

Bibliography

[1] James E. Cutting. Event segmentation and seven types of narrative discontinuity in popular movies. *Acta Psychologica*, 149:69–77, jun 2014. ISSN 00016918. doi: 10.1016/j.actpsy.2014.03.003. URL http://dx.doi.org/10.1016/j.actpsy.2014.03.003http://linkinghub.elsevier.com/retrieve/pii/S000169181400078X.

[2] Robert Gaizauskas and Tarfah Alrashid. SceneML: A proposal for annotating scenes in narrative text. In *Workshop on Interoperable Semantic Annotation (ISA-15)*, page 13, 20