

How can we reconstruct stories based on memories?

Andreea Macovei and Diana Trandabăț

Faculty of Computer Science, “Alexandru Ioan Cuza” University of Iași
General Berthelot, 16, 700483, Iași, Romania
{andreea.gagea, dtrandabat}@info.uaic.ro

Abstract. Memories in literary texts reveal past events which complete the itinerary of a character throughout the book. Whether they are exposed from the point of view of that character or from the perspective of the other characters joining the action, the memories bring novelties about the distant or the recent past to the actual action. This explorative study intends to identify memories, with an ulterior aim of temporally ordering narrative sequences for reconstructing the story of each character. Starting from the premise that memories contain temporal information, the goal is to see if this information can be used to restructure a discourse so that the entire action gets to be chronologically ordered according to the order of the events’ deployment. The described model is tested on a novel of Tash Aw (*Map of the Invisible World*), with possible applications to non-literary texts such as newspaper articles, social media posts, recipes, political texts.

Keywords: Memories, Temporal Information, Text Processing

1 Introduction

Memories are considered sequences of specific events that provide the reader with past images of a character. An analysis of memories seems to be a new subject in the sphere of temporality. Most studies that analyze memories refer to the field of psychology: a past event that is reconstructed in memory can be influenced by construal mindsets [1], while there are some factors that enhance accurate memories such as negative emotions [2].

To our best knowledge, there is no computational analysis that considers the memories expressed in texts and their possible automatic identification. However, specialists have followed several directions when it comes to reordering the discourse. An example could be the model using ordered trees and temporal information where the focus is on variation in narratives in order to sort, to disarrange and to categorize events with zigzag and sylleptic schemes [3].

Also, models trained to identify flashbacks [4] in literary texts demonstrate the increasing concern for identifying the peculiar textual phenomena in literary texts (such as flashbacks and flashforwards, known in narratological terms as analepses and prolepses – past or future events recurring vividly in the mind). Predicting the order of

events can be done with a structured perceptron algorithm [5] characterized by various feature sets and constraints, but this is only possible for texts in which the order of events is aligned with their temporal order (as recipes).

Although an analysis of the events in texts according to the TimeML [6] and ISO TimeML [7] standards can be considered, this proposed method focuses on sequences of events, due to the fact that in the chosen literary text, memories span paragraphs. Further, the event-level delineation will be considered to be able to compare the event level and event sequence levels.

This paper is structured as follow: Section 2 and 3 define memories with examples and particular features, Section 4 presents an experiment done in order to automatically identify memories in literary texts and discusses the obtained results of the experiment. In sections 5 and 6, several conclusions are drawn and the next milestones in this study are indicated.

2 Memories as a distinctive component in discourse

Memories are considered as distinctive temporal segments in a discourse that can be used to reconstruct the temporal axis of a text. They are sequences of events that highlight a certain period in one's life and placing them on an axis of time accurately outlines the route of a character in terms of temporality. The memories are evoked both by the narrator and by the characters involved in the discourse.

Our intent is to identify sequences of memories in their entirety and to place them on the time axis of a character in the book. A possible reordering of events in a book based on the memories displayed evoke more accurately the trajectory of a character starting with the farthest event to which he or she participates and ending with the end of narrative where the last details about his or her person are revealed.

The presence of memories can be extremely obvious (especially if we think of biographies and autobiographies). But, in those types of discourse, the course of action is chronological, and the presented memories only show the events of a personality's life from the beginning to a certain point or until his or her disappearance. And our interest is to use these memories to order the events composing the trajectory of a character in a novel, when they are exposed randomly, but also in cases where a character does not remember anything in a certain period of his or her life, and the memories of the characters he or she came into contact with may reclaim that character's time that he or she does not have the ability to remember.

The selected corpus for our study is a novel by Tash Aw (*Map of the Invisible Word*) which is known for repeated returns in the past and through narratives presented from several perspectives (the narrator's perspective and the perspective of two main characters, the brothers Adam and Johan).

The novel presents the story of the two brothers starting from the period when Adam and Johan were at the orphanage as they were abandoned by their mother at an early age and later, when they were separated and adopted by different families in Indonesia and Malaysia. The period before their abandonment at the orphanage is

evoked by Adam's only memory: he remembers the fact that he was staying with his mother and his brother in a clean house, and was taking care of household chores.

The reader learns about these things while he or she continues to read as the events are not presented in a chronological order and are narrated by the narrator, Adam, Johan and other main characters in parallel. Adam does not remember much of the time he spent in orphanage and with great difficulty, realizes that he has a brother; much later, this determines Adam to go looking for his brother. Adam only remembers the orphanage room, the crowd of children who always surrounded him and that he was not afraid of anything when Johan was near him.

On the other hand, Johan, the older brother, clearly displays memories with Adam during their stay at the orphanage (various conversations with Adam, the moment when John is adopted by his new family and Adam is left in the orphanage, the Christmas spent together at the orphanage, the moments when Adam was sick and John was taking care of him, etc.). In order to have a general view, the figure below (Fig.1) indicates the memories of Adam and Johan during the time spent at the orphanage.

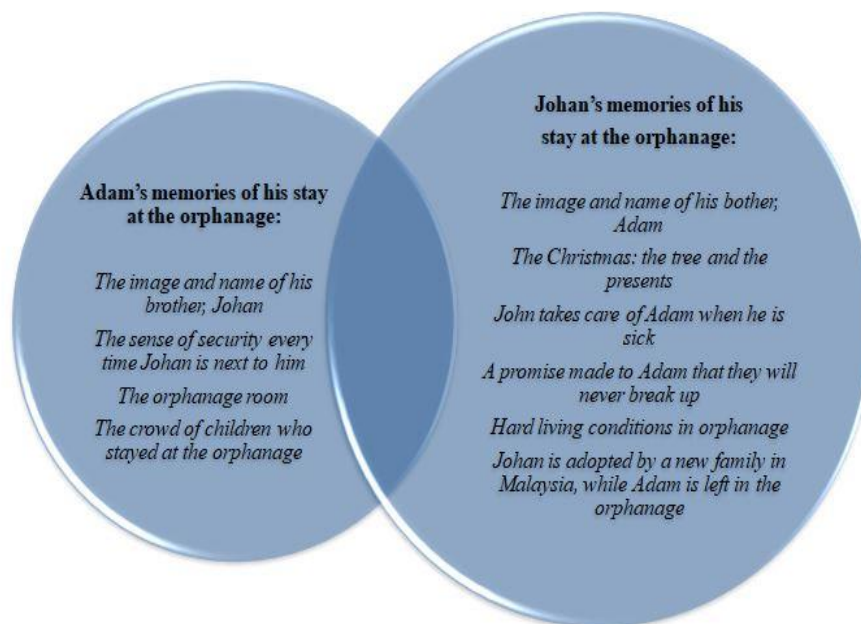


Fig. 1. The memories of the two brothers, Adam and Johan, were exposed from the perspective of everyone while they were in the orphanage.

The intersection of the two circles that include the memories of the two siblings indicates to a certain extent the few similarities between the memories (both remember having a brother and having lived in the orphanage before being adopted by different families). Taking into account Johan's memories, the period during which Adam was at the orphanage can be chronologically highlighted, although his memories are few. Thus, since the memories get to be identified in the text, they have to be placed on a

temporal axis depending on the moment of their action, and must also be centralized so that the perspective from which they are related and the involved characters are the defining aspects of these event sequences.

2.1 Memories in texts

Memories can occur anytime inserted in the text, usually with a clear indication (certain triggers announcing such a sequence), but there are cases when a memory is identified later, without any clue, as in the following example. The memory spans on the second part of the fragment [2], and the first part can be considered as a preamble of the memory [1].

Example 1:

[₁ As Margaret stood in the kitchen waiting for the kettle to boil, she tried to figure out why she had so readily embraced Adam. She could not explain what had made her feel so happy upon seeing him again, why she had been overwhelmed by a sense of relief bordering on joy at finding him safe and eagerly awaiting her return. ₁] [₂ When he stood up to greet her, the expression on his face was one of such hope and anticipation and vulnerability that she had to respond. And so she had gone up to him and put her arms around him, circling his chest tightly, as if she needed to reassure herself that he was still there. She had done so without thinking, and for all the time their hug lasted - three, four, five seconds? More? - it ceased to matter that Din and Mick had been in the room, watching. ₂]

On the other hand, the example below is an evident memory: Margaret, another character in Tash Aw's novel, remembers the story with the small dull buff-coloured bird attacking the snake [2]. The story is introduced by the verb *remembered* [1]:

Example 2:

[₁ She remembered ₁] [₂ something she had seen when she was a child on Irian: a small dull buff-coloured bird attacking a snake, falling on its thick coils in swift stabbing dives until the reptile was forced to retreat. A villager told her that the bird was a mother protecting her nest from predators, to which Margaret had said, Oh, I understand, the bird loves its children so much that it becomes brave enough to attack a much bigger animal. And the villager said, No, it is not love that makes the bird do what it is doing. It is foolishness. This bird actually believes it is stronger than the snake, it actually believes that no harm can come to it. It is a very silly bird. ₂]

Memories are part of a character's journey, as other past details of the character can be revealed and cataloged as events or sequence of events. This character's trajectory along a literary text is defined as a time track [8] and aims to capture all the sequences of events (named *time segments*) in which that character is involved. In this first phase, event sequences are taken into account and not the events, as well as the temporal relationships between these sequences, locations, and characters that help in determining time tracks.

The time segments are classified in five types [9]: *Narration* (a sequence of events), *Remembers* (memories that describe certain events that have taken place in the past), *Supposition* (the assumptions that a character can imagine), *General Knowledge* (exact information and knowledge noticeable in the text), and *Fiction* (another fiction that appears in the text defined as a story within a story).

The ultimate goal is to identify these *Remembers* type sequences depending on the clues found at the text level, so that the time tracks based on all types of sequences can be represented graphically in the form of time diagrams.

3 Features of memories

With regard to the features of the memories identified so far, it can be observed that a memory is based on the trigger. Most of the times, this trigger is the verb *to remember* and other synonymous verbs, and few are the cases where this trigger is missing. Such a case could be the Example 1, where the first phrases [1] reveal the introductory part of the memory, and the actual memory is surprised in the next part [2]. After this trigger, the memory is evoked and if there are, some effects of that memory that occur at present. Fig. 2 emphasizes a memory with such a structure:

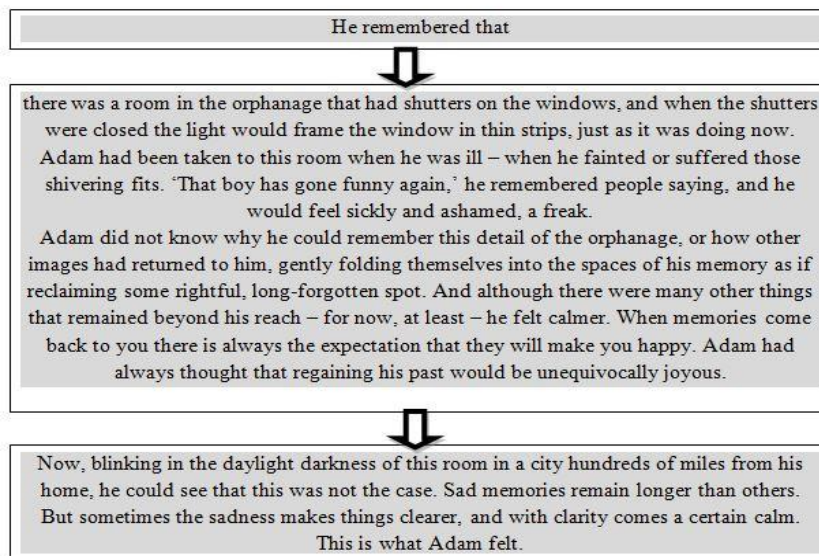


Fig. 2. The structure of a memory: trigger, memory, effect.

Another important aspect of these memories represents the variety of their construction. Unlike the memories that clearly illustrate the past of characters with real events, there are cases where things are not exactly the same. So, starting from a real fact, a character has the ability to remember a fiction such as the song of a bird, an object, a poem, or a novel he or she has read before. Such memories can hardly be placed on an axis of time since they cannot be compared with other past or present events (as in Example 2). Also, the following example attests to this, as Johan remembers the Christmas tree and this memory can be somehow correlated with the period in which he was in the orphanage:

Example 3:

Johan thought of the Christmas tree they'd had at the orphanage, which never had any lights on it, and wasn't even a proper Christmas tree, just a half-dead sapling with its branches cut to make it less messy. Sometimes there would be a few trinkets hanging from it, palm leaves woven into balls, or some fruit on short lengths of string, but it didn't matter, because none of the children knew what Christmas was. There were presents, like an old pair of socks or box of biscuits or a toy that some rich city kid did not want any more but you did not know better, and when you do not know better you are happy for what you have.

There are also situations where a sensation, a dream, an imagination (something unreal) triggers a real memory, but this time, the reader will be able to delimit it in time. This time, Johan remembers the holiday spent with his new family at the seaside and it all goes from the drops of rain which he has the impression that he hears them falling on sand:

Example 4:

It sounded to him like drops of rain falling heavily on sand, the start of a storm at the seaside. He remembered his first holiday with his new family in Port Dickson.

4 Identification experiment of memories

In order to implement a program to identify memories from the texts, the Weka tool [10] is used for a first stage of training and testing. Using the sequences of memories already annotated from the Tash Aw's novel, the next step was to extract some features of these text sequences that make them different than other sequences.

For the beginning, features that referred to the form are considered. Those features are:

- the presence of the trigger verb *to remember* or other synonymous verbs (*Feature 1 – F1*);
- the quote marks (when the memory is related with the exact characters' words – *F2*);

- the indirect speech (if quotes appear – *F3*);
- the verbs at past tense of phrases (the variation of the verb tenses is much more visible in Romanian language than in English; often, specific times such as past perfect tense and past tense are used for exposing the memories – *F4*).

After testing and training on a relatively small corpus (several sentences) at phrase level, the results were small but rather encouraging (40% for precision and 50% for recall with Naive Bayes classifier). The results reflect the very few features, but also the length of the corpus. On the other hand, the past tense of verbs in memories is not a particular feature, since the events in text are exposed using past tenses of verbs.

It is also difficult to stop at the phrase level, since in the literary text, the memories span on several phrases; most of the time, a memory covers an entire paragraph. For this reason, attention was also directed to content features. These content features refer to:

- the number of characters found in a specific paragraph (the appearance or disappearance of one or more characters in the previous – *F5* – and next paragraphs may indicate a memory – *F6*);
- the existence of a temporal expression (it may indicate a rupture between the past and the present action – *F7*);
- but also the timeline match between the paragraphs (the action in the paragraph is not presented in the chronological manner, as it happens in the previous – *F8* – or next paragraphs – *F9*).

4.1 Results of Weka model

For testing, both the features of the form and the content (nine features) are considered in a first phase. Then, a few features have been eliminated to find out if there are some relevant and less relevant features of memories. In Table 1, the best results obtained after running Weka are shown.

Table 1. Results in Weka

	Classifiers			
	Naïve Bayes		J48	
Tested features	Precision	Recall	Precision	Recall
All features	76%	78%	60%	71%
F1-F5-F6-F8-F9	60%	71%	58%	65%
F1-F2-F3-F4-F7	51%	56%	43%	48%

As can be seen, results are higher than those originally generated. The extensive range of features has helped to achieve better percentages for precision and recall. In the running process, the percentage of memories that can be found in a novel (the percentage for Tash Aw's novel is about 10%) was also taken into account, which means that only 10% of the training and testing corpus (20 paragraphs) was made up

of memories. Also, the results are higher with the Naive Bayes classifier as the test model is small, which is a disadvantage for decision trees' classifiers.

What is interesting to observe is that content features seem to be defining for memories without omitting F1, the feature that refers to the verb trigger to remember. This can be simply explained by the fact that F2 and F3 are not so often seen in memories (quote marks and indirect speech), unlike F5 and F6, but also F8 and F9 (the character who remembers something brings to mind the happenings where other characters have taken place or just a few of those mentioned before the memory).

The same is true for the time line match (F8 and F9): the reader perceives faster that the memory is part of another period of time that does not match the one in the next or previous paragraphs and this is also available for Example 5, where [1] is the previous paragraph (or the previous event sequence) before the memory's one:

Example 5:

[1 He opened the refrigerator and found a hard, putrid lump. Its label read ROMANO WHEEL. It did not smell at all appealing, but memories of dinners at home had sharpened his hunger and he cut a wedge of cheese and ate it. He cut the rest of the lump into pieces and put them on a plate, wanting to arrange them like spokes of a wheel but finding instead that they created a miniature atoll, like the haphazard clumps of coral on the seabed as he drifted over them. my father was wearing a blue shirt when they took him away it was definitely soldiers and not the PKI. As he ate, he opened his notebook and wrote:

THINGS TO TELL MARGARET

my father was wearing a blue shirt when they took him away it was definitely soldiers and not the PKI.

(Communists always kill their captives,
soldiers don't always)

(And I can also recognise communists -
they dress very badly because they are poor)

my father's only crime is to have been born
a foreigner

i searched the house for clues but found
only pictures of you and my father

you looked very happy ₁]

[2 He recalled scrabbling around nervously among Karl's papers, afraid he was transgressing some undefined boundary. The memory of Karl's room - his small desk, his curling handwriting on sheets of paper, the smell of mothballs and old books - was clear and unwavering. ₂]

These results are influenced by the situations in which memories occur in a conversation between two or more characters, one of whom gets to say exactly what he or she remembers after another asked him or her to do so. Example 6 reveals an error case:

Example 6:

Farah turned around to check that Bob was asleep. What comes back to you?

The orphanage. The boys. The rain dripping through the roof, the sound it made, like a ticking of a huge clock that would never stop. The dorm was long and thin and there were rows of camp beds, just a sheet of canvas between pieces of wood, not even a proper bed, and the rows were so close together there was barely any space to walk between them. You could hear the breathing of the other boys when they were asleep and every night there would be someone crying in his sleep.

5 Further work

There is a lot of work from now on. The next step would be to train and test the created model on the whole corpus to get the most accurate results. In the future, it is desirable to implement a program that identifies all the memories in the texts, based on the already found features. Once a larger corpus with memories is developed, one can take into account a more specific classification of them as well as varied traits.

The next step is to test the model and the found features on other types of text (non-literary discourses) such as newspaper articles, social networking posts (blogs), culinary recipe texts, political discourses, etc. In these texts, it is expected that the announcement of a memory should be obvious, the evocation of that remembered event with a shorter length than in literary texts as shown in the example below taken from a culinary recipes site¹:

Example 7:

You can use any coconut oil or butter if you do not prefer lard. If you put something sweet in it [bread], you have a cake, as there is so much flour.

I have used mineral water as I have remembered that this was how mom had made the cake doughs sometimes.

Unlike literary texts, in non-literary discourses, memories are easy to find (the verb *remember* is the clue word) and often, bring additional information about a present action, as in the above example. However, this model needs to be improved by looking for other features that make the memories worthwhile, and an implementation of it is necessary to lay the foundations for a tool to identify memories from written texts based on the features already identified.

Then, we intend to extract the events of our corpus and the relations between them. For this step, we have searched some instruments that may help us and the Fred tool [11] seemed a very useful tool in order to identify the memories of the novel as events

¹ The example is taken and translated from the website: <https://dietaderotatie.ro/paine-vegana-fara-gluten-cu-faina-de-orez-soriz-tapioca/>.

as this tool is able to automatically produce RDF / OWL ontologies and linked data from sentences.

Those ontologies offer us information regarding the event of the phrase and its connections with other components of the phrase. Fig. 3 is the graphical representation of the sentence of *Example 6* (*What comes back to you?*), an introductory sentence of a memory, using Fred tool, where the verb *comes back to* is classified as an event of that clause.

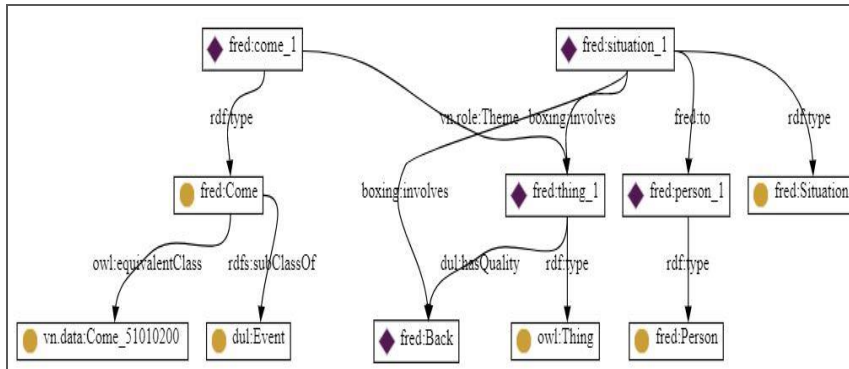


Fig.3. Representation using Fred tool of an introductory sentence of a memory

Also, placing a memory on a time axis would allow any reader an overview of the evolution of the entire character. As a long-term goal, a visual representation that includes the events and memories of a character (related from their own perspective or that of the other characters with who he or she came into contact with) and the temporal connections between them will be implemented; this will allow a clearer view of the whole action, especially for long texts such as belletristic, where temporal references may not appear, but returns in the present and past are very frequent.

6 Conclusions

This study seeks to delineate memories from a text to better place all actions in a temporal context. It also raises the premise that relying on the memories of a character, stories or certain periods of his or her life can be rebuilt. Thus, memories bring new information about the positioning of the character in relation to oneself or other characters in the past that may or may not influence the present.

Memories are part of a character's life, and can reconstitute natural trajectories of the past that mark the evolution of that character. Besides the memories of a character about himself or herself, and the memories of the characters he or she came into contact with, bring information about his or her trajectory all along the book. For example, if there is an actor of action who does not remember anything in a certain period of his or her life, the memories of the other actors are helpful in order to reconstitute and put together events that have happened long ago.

This is especially true in texts where a narrator wants to capture the readers' attention, using all sorts of narrative techniques to change the natural order of actions. In this way, the reader finds himself forced to be more careful to reorder chronologically those events in his or her mind. Memories can change the course of the stories in the sense that they come with additional aspects to cover the gaps of certain periods in a temporal context. This analysis can allow a modeling of the whole discourse as it has the premise of ordering events as they have happened.

Identifying memories in texts provides the reader an interactive way to return to some aspects of childhood, adolescence, youth and other periods of the character's past not mentioned so far. This can be done without having to go back to reading, but in an automatic way.

Acknowledgment

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS/CCCDI – UEFISCDI, project number PN-III-P2-2.1-BG-2016-0390, within PNCDI III.

Extracts from *MAP OF THE INVISIBLE WORLD* by Tash Aw, Copyright © Tash Aw, were used by permission of The Wylie Agency (UK) Limited. The Humanitas Fiction Publishing House, Bucharest, has kindly accepted to offer us the Romanian version of this novel, translated by Florica Sincu, for research purposes.

References

1. Kyung, E. J., Menon, G., Trope, Y.: Reconstruction of things past: Why do some memories feel so close and others so far away? *Journal of Experimental Social Psychology*, 46(1), 217-220 (2010).
2. Kensinger, E. A.: Negative emotion enhances memory accuracy: Behavioral and neuroimaging evidence. *Current Directions in Psychological Science*, 16(4), 213-218 (2007).
3. Montfort, N.: Ordering events in interactive fiction narratives. In: Magerko, Brian S., Riedl, Mark O. (eds.) *Proceedings of the AAAI Fall Symposium on Intelligent Narrative Technologies*, 2007, pp. 87-94 (2007).
4. Wu, H. Y., Young, M., Christie, M. A.: Cognitive-Based Model of Flashbacks for Computational Narratives. In: Sturtevant, N., Magerko, B. (eds.) *Proceedings of Artificial Intelligence and Interactive Digital Entertainment* (2016).
5. Abend, O., Cohen, S. B., Steedman, M.: Lexical Event Ordering with an Edge-Factored Model. In *HLT-NAACL*, pp. 1161-1171 (2015, June).
6. Pustejovsky, J., Castano, J. M., Ingria, R., Sauri, R., Gaizauskas, R. J., Setzer, A., Radev, D. R. et al.: TimeML: Robust specification of event and temporal expressions in text. *New directions in question answering*, 3, 28-34 (2003).
7. Pustejovsky, J., Lee, K., Bunt, H., Romary, L.: ISO-TimeML: An International Standard for Semantic Annotation. In *LREC*, Vol. 10, pp. 394-397, (2010, May).
8. Cristea, D., Macovei, A.: Why time resembles rail yards? A way to look at time in text books. *Romanian Journal of Information Science and Technology*, 19(1-2), 31-43, 2016.

9. Macovei, A., Cristea, D.: Time Frames: Rethinking the Way to Look at Texts. In: Bunt, H. (editor) Proceedings of isa-12, the Twelfth Workshop on Interoperable Semantic Annotation in conjunction with LREC 2016, Portoroz, Slovenia, pp. 59-62, 2016.
10. Frank, E., Hall, M.A., Witten, I.H.: The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, Fourth Edition, 2016.
11. Gangemi, A., Presutti, V., Reforgiato Recupero, D., Nuzzolese, A. G., Draicchio, F., Mongiovi, M.: Semantic web machine reading with FRED. *Semantic Web*, (Preprint), 1-2, 2016.