Visual Exploratory Tool for Storyline Generation

Erk Ediz Akyiğit
Sabanci University

Tuğkan Cengiz
Sabanci University

Onur Burak Yıldırım
Sabanci University

Selim Balcisoy*
Sabanci University

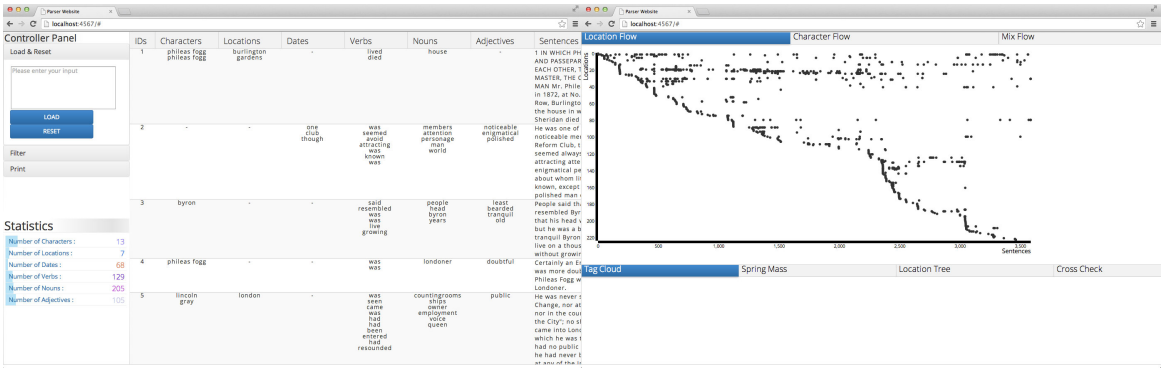

Figure 1: Overview of the proposed tool.

1 Introduction

Storyline visualizations are useful to let people explore works of literature. It was first introduced in XKCD's Movie Narrative Charts as a hand-drawn illustration [1]. Tanahashi and Ma proposed some considerations on the design of the storyline visualization and automation became meaningful because of its aesthetic and clear representation style [2]. Shixia Liu and Yingcai Wu's StoryFlow had become the next step in storyline visualization [3]. With the efficient optimization approach on automation, representation of the complex stories became efficient and enabled users to track and understand the story easier. In all cases before generating the visualizations, an input file must be created manually. In this work we propose a semi-automatic exploratory visual tool for preparation of such input files.

The tool first preprocesses a novel by separating time, space and character components according to an identifier and then classifies them. After preprocessing, the tool allows users to make modifications of location, character and storyline with visual interfaces. Our tool leads user to efficiently analyze whole storyline with character and locations in a continuum provided by the story.

2 Interface

The tool consists of seven different modules divided into two different screens as presented in Figure 1.

The left screen contains a control panel and manuscript view where the extracted data from the novel is presented to users as the representation of the database. The control panel works as a search and filter mechanism to simplify analysis of the original text of the novel. Considering the detail level of the novel this kind of representation is essential, because at some point users have to dig into the raw data to investigate unclear sections of a novel.

The right screen contains five different modules for visual investigations. A node link diagram expresses the strength of relationships between characters, where every character is represented as nodes and lines amongst them shows how strong the relation between them is. It searches character duple in the novel and increments the count where these characters are encountered in the same sentence.

The second one is tag cloud and aims to show the context of the relation between two characters by presenting the words that are used between them in the novel.

A cross check module is there to show possible mistakes originating from Named-Entity Recognition Process (NER) by displaying words that are tagged multiple times by different tags.

The next module is a tree representation of locations occurring in the novel, as in the novel there are many different location indicators such as countries; cities etc. and these locations cannot be put in an order automated since there are many places with same name in different locations.

The final module contains storyline line graphs, where the location, character or combination of both is represented on the Y-axis and the X-axis represents the sentence number. A cross examination of story flow with locations, characters and their combination views, the user can understand who disappeared at which location in which sentence or other changes in the story flow. The location graph contains the locations at Y coordinate and the black dots show the locations which are told by book. The character graph contains the same structure as the location flow but with the characters. The combination graph presents the combination of these two graphs where dots represents the characters with the locations expressed as different colors.

3 Preprocessing

Preprocessing phase of the project is based on NER, which is the task of information extraction in order to classify elements in a text. It contains three steps: Parsing, Database Management and Relational Matrix Construction. The first step is splitting the

* balcisoy@sabanciuniv.edu

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

whole novel into sentences and extraction the components of these sentences by using the parser: TREAT [4].
Second step is storing these named-entities in the database by keeping all chapters in different tables for differentiation of progress of events according to flow of the novel. Each row in a table corresponds to a sentence within that chapter respectively and characters, locations, time, adjectives, nouns, verbs and the sentence's itself is kept at the corresponding columns.
The last step is the organization of the relations between characters by creating a relational matrix amongst characters. The NxN matrix, where N is the number of characters, generated in this process by incrementing the corresponding row and column value in the relational matrix for each character duple in each sentence for the whole novel.

4 Case: Around the World in Eighty Days

We analyzed Jules Verne's adventure novel, 'Around the World in Eighty Days' as a case study, as the plot takes place in many different places and large group of different characters take part in the story.
The parser splits the novel into 37 chapters and the TREAT library fills the database tables. A representation of the database projected to manuscript view and the cross check module reports tagging issues.
For example, Aouda is both tagged as a location and a character by the parser but when we analyzed a few of the sentences that Aouda is being mentioned using the manuscript view tool, we could easily identify that the word Aouda is representing one of the leading characters in the novel. Also, the tool tagged Passepartout, the servant of Phileas Fogg, as a noun or adjective in most of the sentences. After finishing the corrections, we can start to create the location tree. At each modification performed with the tool, the program manipulates the storyline flow graphs' location information.
By scanning through the storyline flow (Figure 2 labeled as Mix Flow) , we can see the characters and places on a single time line. Also by examining the node link diagram and tag clouds generated by the relational matrix, we can understand the context of the relationship between these characters in order to make sense of the events taking place. The example below shows a sample process for instrumentation of the relational matrix:
"Phileas Fogg lost no time in going on board the Carnatic, where he learned, to Aouda's great delight—and perhaps to his own, though he betrayed no emotion—that Passepartout, a Frenchman, had really arrived on her the day before."
In the first run, the tool recognizes Phileas Fogg as the first character and increments corresponding columns of Aouda and Passepartout by one. In the next run, it recognizes Aouda and increments the Passepartout's column. In the third run it fails finding another recognized character other than itself and terminates the loop.

5 Conclusion

We presented a tool offering a new way of analyzing a novel in terms of storyline, characters, locations and taking consideration of relations between all of them. It can save significant amount of time for creating editorial and narrative works such as the input data for storyline visualizations or reviewing a book draft. In future we will be improving our tool such that error rate of the tools findings will be decreased and texts other than novels with linear time, can be parsed and presented.

References

[1] 'Movie Narrative Charts' 10 March 2014 [Online]. Available: http://www.explainxkcd.com/wiki/

[2] Tanahashi Y., Ma Kwan-Liu, 'Design Considerations for Optimizing Storyline Visualizations' IEEE Transactions on Visualication and Computer Graphics 18(12): 2679 – 2688, Dec. 2012

[3] Shixia Liu and Yingcai Wu and Enxun Wei and Mengchen Liu and Yang Liu, 'StoryFlow: Tracking the Evolution of Stories', IEEE IEEE Transactions on Visualication and Computer Graphics 19(12): 2436-2445, Dec. 2013

[4] 'TREAT' 10 March 2013 [Online] Available: https://github.com/louismullie/treat/wiki/Manual

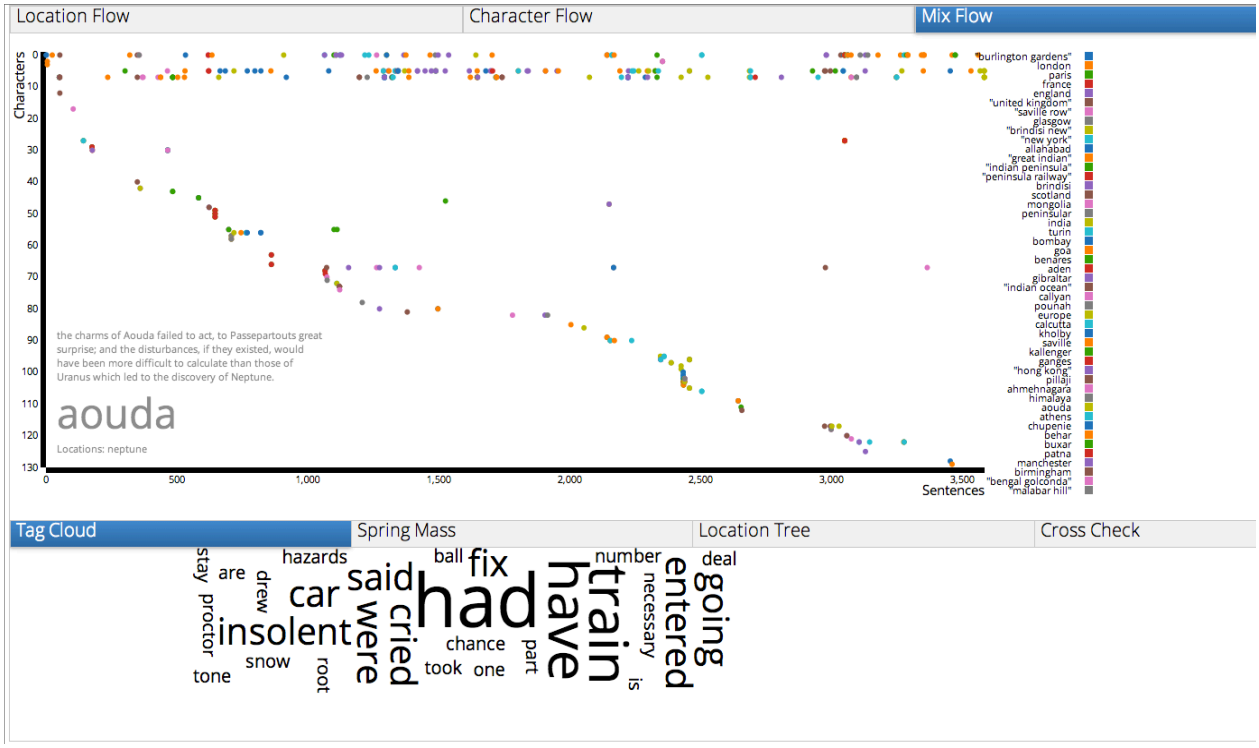

Figure 2: Storyline of the novel Eighty Days around the world.