Master's Thesis

ANALYZING USER INTERACTION LOGS OF AN EDUCATIONAL VISUALIZATION SYSTEM TO UNDERSTAND HOW STUDENTS GENERATE INSIGHTS

Kuatbek Mukabak

Department of Computer Science and Engineering

Graduate School of UNIST

2018

ANALYZING USER INTERACTION LOGS OF AN EDUCATIONAL VISUALIZATION SYSTEM TO UNDERSTAND HOW STUDENTS GENERATE INSIGHTS

Kuatbek Mukabak

Department of Computer Science and Engineering

Graduate School of UNIST

Analyzing User Interaction Logs of An Educational Visualization System to Understand How Students Generate Insights

A thesis

submitted to the Graduate School of UNIST

in partial fulfillment of the

requirements for the degree of

Master of Science

Kuatbek Mukabak

06/11/2018 Month/Day/Year of submission

Approved by

_____

Advisor

Sungahn Ko

Analyzing User Interaction Logs of An Educational Visualization System to Understand How Students Generate Insights

Kuatbek Mukabak

This certifies that the thesis of Kuatbek Mukabak is approved.

06/11/2018

signature

_____

Advisor: Sungahn Ko

signature

_____

Jaesik Choi: Thesis Committee Member #1

signature

_____

Young-Woo Park: Thesis Committee Member #2

Abstract

Visual analytics systems have been becoming popular in many domains. Recently, a visual analytical tool, VAiRoma is designed in educational domain to support students learn the history class. However, how users are interacting with such systems is still not known enough. In an educational domain, it is important to know how users are gaining insights. It may give us an opportunity to understand the user's learning style, so that we can design better visualization tools in the future. In this thesis, I will analyze the interaction logs of an educational visualization system, VAiRoma, in order to explore how users generating insights via the system. Based on the results, users tried more explorative interactions at the initial stages of their insight generation path. In the middle of the path, users mostly read some textual information. Toward the end, they attempted to show their understandings from what they learnt by creating an annotation. There is also a cyclic behavior of an insight generation path. In 38% of cases, during the annotation creation process, the users cancelled to "create an annotation" and went back to read some textual information.

Table of Contents

List of Figures

List of Tables

I. Introduction

The emerging popularity of visual analytics influenced researchers to apply it in a wide range of domains. There are examples of visual analytic tools used in medicine [1], traffic analysis [2] [3], cyber security [4], online shopping [5] and education [6]. Although, many researchers have been trying to propagate the power of visual analytics in various domains, how users are interacting with such systems is still not known enough. Similar research has been conducted in specific domains such as commercial websites to understand clients and their needs [7]. However, to my best knowledge, there was no research on user interactions of educational visualization systems. Although, like other visual analytic system, one of the primary goals of educational visualization systems is supporting users to gain insights, the insight generation process in learning is complex and requires deep understanding of user behavior. The users of such systems are characterized by using the interface not only for seeking simple facts and short answers, but also for trying to understand notions, ideas and concepts to make own knowledge. By studying the user interaction patterns of such systems, it is possible to adapt the learning environments according to user behavior and natural characteristics of human perception.

This thesis represents an attempt toward studying this problem. For this, I used a web-based visual analytics system called VAiRoma [6], which is designed to support users in learning Roman history. For this experiment, the dataset of the system is changed from Roman history to contemporary history of the world. The experiment is conducted in a history class for 62 students during one semester period. The students used the VAiRoma system to learn their subjects and submit their homework. The homework was to do an exploration around a selected topic from a topic list provided in the system (there are 40 topics in the system, each topic consists of a set of related keywords such as "world war", "Soviet Union") and write an annotation at the end of the exploration. All the user interactions (clicking a tab, scrolling a window or zooming a map, etc.) are logged in the server. I used these interaction logs data for my analysis to reveal (hidden) patterns in user behaviors, understand insight generation processes and identify possible requirements for designing educational visualization systems.

To see the high-level view of an insight generation process in our system, I picked several important interactions (Wikipedia article reading and annotation related interactions) and analyzed the user interaction sequences in terms of those interactions. As a result, it was revealed that the insight generation process is an iterative process. During the annotation creation process, 38% of the paths come back to the Wikipedia article reading state several times to gain more information or to obtain some missing information. And even after the annotation is created successfully, 16% of annotations are edited.

In general, students tried more explorative interactions for their initial steps of an insight generation path. It includes searching for a Wikipedia article, playing with a map and other visualizations in the system. Then, for the main part of the insight generation path, they mostly read the Wikipedia articles, searched for a new Wikipedia article, read it or read the already visited Wikipedia article. This behavior is repeated for several times. To the end of the path, they mostly read the article, tried to create an annotation, but "cancelled to create" to gain more information.

Here is the main contribution of this project:

1) I identify an insight generation path in such systems;

2) I analyze user behaviors to find key patterns;

3) I discuss design recommendations for educational visualization systems based on the results.

II. Background and Related Works

Researchers tried to analyze interaction logs in visualization systems from various aspects to find common patterns [8], to study user behavior [9] [10] or searching for insight provenance [11] [12] [13]. Han et al. summarized these interaction analysis tasks and categorized them into four types - assess usability, assess utility, learn about users and understand usage patterns [14]. Assessing usability and utility might be evaluated by quantitative analysis, whereas learning about users and recognizing usage patterns may require deep understanding of user behavior, or sometimes even deeper, the cognitive knowledge generation process in human mind. Despite the complexity of human behavior on visualization systems, there are plethora of works dedicated to study this field.

Among common user pattern related works, Dabek and Caban tried to model user interactions with a purpose of guiding a user through a visual exploration process [9]. They proposed a grammar-based model, which can learn from user interactions. By detecting common patterns among subjects, it generates some rules, which then can be used as a suggestion tool for new users in a visual exploration path. Another similar tool EventAction [10] also learns from user interactions and generates suggestions for temporal event sequences. Common patterns among users also can be used for user clustering. Wang et al. presented a practical and scalable clickstream tool for user behavior analysis [15]. It clusters the users by partitioning a similarity graph. The visualization of this algorithm clearly distinguishes main clusters among users and the hierarchical relationship between some clusters. Although these works are dedicated to study user behaviors, they mostly focus on low level, physical interactions of users rather than understanding provenance of human decision making.

Recently, there are some attempts toward modelling user behavior and understanding how people interact with visualization systems [12] [11] [13] [16]. User behavior in visual exploration path can be modelled as Markov chain process with transitions between interaction, mental and computational states [16]. This model describes the insight generation process more accurate, since it includes both cognitive and computational states, which are essential for insight gaining in exploratory visual analysis. Another cognitive model is proposed by Chen et al., where human decision making through interaction is modelled as Partially Observable Markov Decision Process (POMDP) [11]. They tried to predict human behavior by solving the POMDP using machine learning and achieved a good result. According to some of these research works, there is a higher probability, that some of the cognitive processes may occur together than others [13] and some interaction patterns inclined to led to insights [12].

Other research works contributed to help for analyzing insight provenance by suggesting fundamental taxonomies of an insight generation path [17] [18] [19]. Gotz and Zhou introduced "Action", as a semantic building block for insight provenance [17]. It is used to describe atomic analytic step performed by a user. They also proposed 21 action taxonomies with 3 top-level categories: exploration actions, insight action and meta actions. Another taxonomy is suggested for more higher-level analysis with only 7 categories: select, explore, reconfigure, encode, abstract/elaborate, filter, and connect [19].

III. Data Collection

In this section, I will discuss about the experiment settings and data collection processes. The experiment is conducted in one semester period for a history class by iVADER lab. The contribution of this thesis is mainly analyzing the collected data. However, it might be essential to describe the experiment settings to understand the data analysis processes in detail.

3.1 Experiment Settings

For the data collection process, I used VAiRoma (*Figure 1*), a web-based educational visual analytic tool designed to support the users to learn the Roman history by analyzing 189,000 Wikipedia articles [6]. For this experiment, the dataset of the system is changed from Roman history to contemporary history of the world. There are given 40 topics in topic view, where each topic consists of a set of related keywords (*Figure 1* - A) and the users are asked to choose one topic and perform an exploration around that topic using the VAiRoma system. At the end of the exploration, the users should leave an annotation from what they learnt. The total number of participants is 62. The experiment is divided into five different periods, with 5 annotations in each period. Therefore, in an ideal case, each user should leave 5x5 = 25 annotations, and 1550 annotations from 62 users. However, the users created 1189 annotations

in total. All the user interactions are logged into the server, including some low-level interactions such as dragging the map or scrolling a view.

3.2 The VAiRoma System

The VAiRoma system filters Wikipedia articles based on selected topic (*Figure 1* - A) and time (*Figure 1* - C), and visualizes them as a heat map (*Figure 1* - B). View "E" lists the Wikipedia article titles and view "D" shows the location names indicated in the Wikipedia articles. The user can click one of the article titles to read the article. The article content will be available in wiki view in "G". There are also other views (*Figure 1* -G) such as table view – a detailed description of relationships between articles, locations and topics; annotation view – the list of created annotations by the user; group annotation view – annotations of other members in the group; and keyword view – searched keywords by group members. Furthermore, users can search for the Wikipedia article from the search box (*Figure 1* - F).
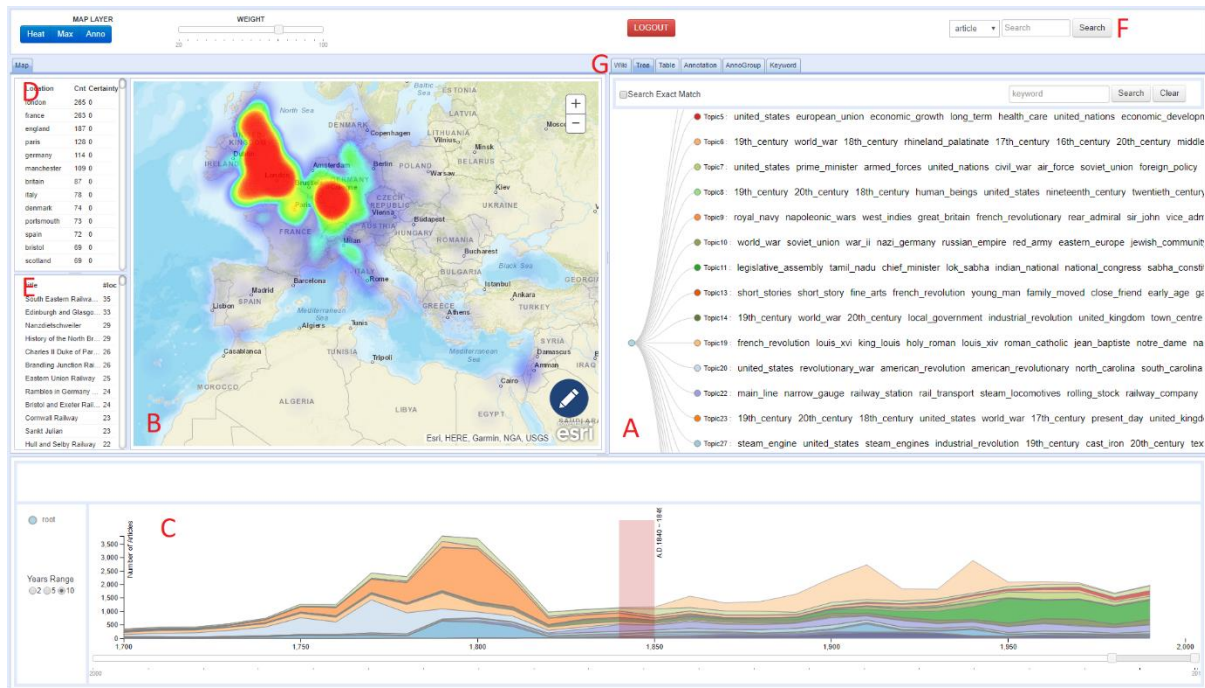
Figure 1: VAiRoma - visual analytic tool for learning Roman history

3.3 Log Format

An example of logged interactions is given in Table 1. Each row shows one logged interaction with information about "user" – the name of the user, "cycle" – in which cycle the interaction is performed, "session" – each login interaction creates a new session, but the session value is independent between cycles, "view" – in which view the interaction is performed, "action" – the name of the interaction, "detail" – more information about the interaction, "time" – the timestamp when the interaction is

happened. After the data pre-processing steps, the total number of logged interactions is 42,935, with 138 unique interactions.

user	cycle	session	view	action	detail	time
***	1	1	login form	logged in	, window size : 1366 * 638, screen size : 1366 * 768	1505285408408
***	1	1	right tab	changed to AnnoGroup		1505285414735
***	1	1	right tab	changed to Keyword		1505285417592
***	1	1	right tab	changed to AnnoGroup		1505288870034
***	1	1	map0 view - title table	sorted	by title	1505288874206
***	1	2	login form	logged in	, window size : 1366 * 638, screen size : 1366 * 768	1505288922973
***	1	2	map0 view	navigated	0, -7704378.646, 10454613.066000002, -4964645.035, 7793612.073000001	1505288929582
***	1	2	topic tree view	navigated	1.0352649238413776, 10.463232191876841, -684.4587935399917	1505288938528
***	1	2	main view splitter	changed width	left panel: 780, right panel:571	1505288958990

Table 1: The logged interactions

IV. Approaches for Insight Generation Path

In this section, I will explain my approaches to reveal an insight generation path. Before analyzing the user interaction logs, it is necessary to understand the notion of the insight generation path. Gotz and Zhou [17] described an insight provenance as "a historical record of the process and rationale by which an insight is derived during a visual analytic task". We can refer to this definition to define insight provenance in our system. As I mentioned in earlier section, users explored the system around one topic and created an annotation. Therefore, when the users created an annotation, probably they obtained an insight and the insight provenance in our system is the path which leads to annotation creation.

4.1 Filtering the Paths

Before starting the analyses, the interaction logs are filtered by some rules. Firstly, the user logs are chopped by "created the annotation" interaction, and I refer each chopped a sequence of interactions to a "path". After chopping, I obtained 1,189 paths. The goal is to discover main interactions and interaction sequences by analyzing these paths which could be contributing to insight generation. Because, there might be some noisy interactions, which have no effect for knowledge generation process. *Figure 2* shows two annotation creation process types. The first type (Type A) is performing

full elements of an insight generation process, and another type (Type B) is creating an annotation without any exploration (i.e. there is no information about the provenance of the insight). As a result, I removed all Type B paths and kept only the Type A paths for my analysis.
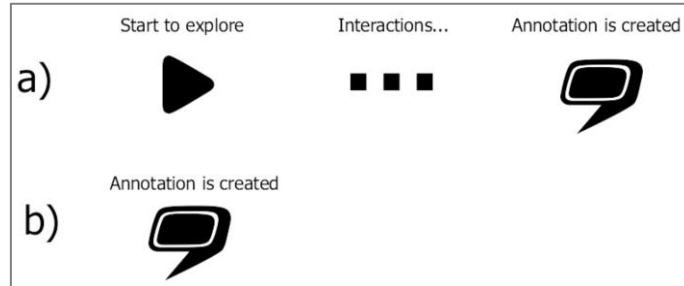

Figure 2: Overall architecture of a detected annotation creation process

To see an overview of the paths, I first created a histogram of number of interactions in each path (Figure 3). In order to enhance the readability of the graph, I draw only 1,120 paths out of 1189, where each path consists of less than 100 interactions. The histogram shows that 221 annotations are created within 1 step which means they only have "created the annotation" interaction. Since those paths do not have any exploration related interactions, it is impossible to obtain the origin of insight. Therefore, I defined a rule to get insight generation paths. The rule is a path should contain at least one source of insight. In the visualization system, there are only three possibilities that the user can read a text: wiki view, tabular view and group annotation view. A user can read a Wikipedia article from wiki view and the tabular view or he may read other users annotations from the group annotation view. Since the user creates an annotation at the end of the exploration path, I hypothesize that there should be some meaningful information for the user during the exploration.
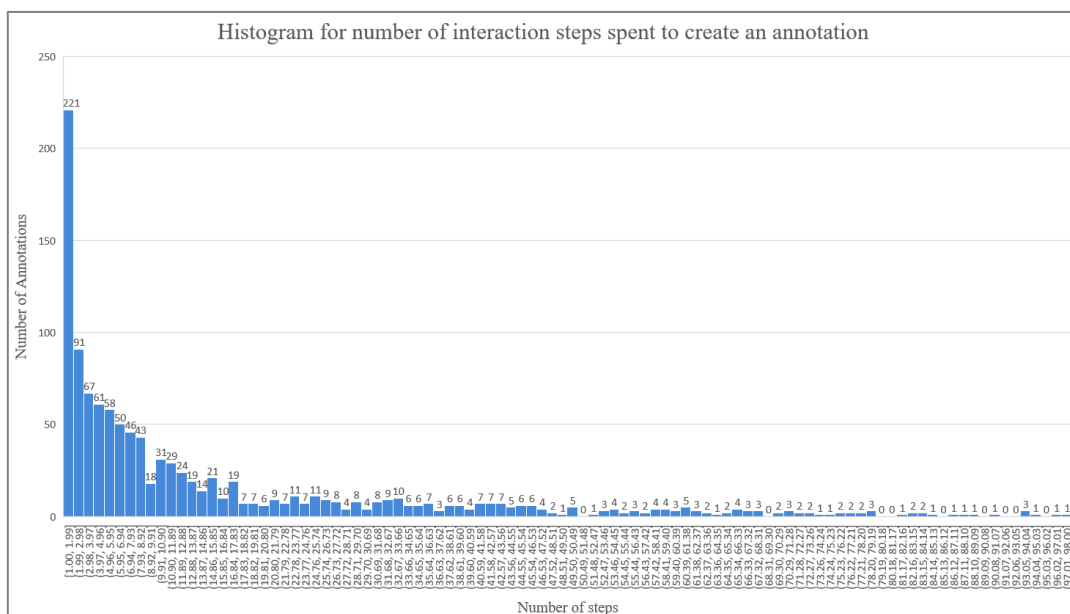

Figure 3: Histogram of number of interactions in each path

By applying the rule, I collect 813 paths (Figure 4). To achieve more accurate results, the three sources of insights (i.e. wiki view, tabular view and group annotation view) require separate analysis. Since the wiki view is used frequently comparing to other two sources of insights, I focused to wiki view and keep only 560 paths, where users read some Wikipedia articles. One interesting pattern is that some of the paths include Wikipedia article reading interaction, but without opening the article in that path. It indicates that the Wikipedia article was opened before starting the exploration path. Since it is important to know that how an article is found by the user, I further removed paths that do not have article opening interactions. This final cleaning step left 546 paths for my analysis. These paths are used for the rest of this paper when I refer to insight generation paths.
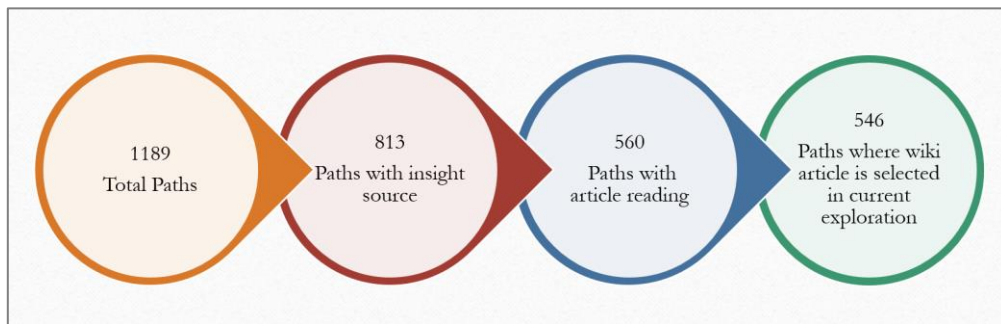

Figure 4: The number of paths left to analyze after each cleaning step

4.2 The High-Level View of an Insight Generation Path

The high-level view of the insight generation path corresponds to observing influence of article selection and article reading for the annotation creation process. Since I have 138 unique interactions, it is almost impossible to visualize all the interactions and find some common pattern among the paths. Therefore, I selected some of the important interactions in an annotation creation process and visualized them to see the relationship among these interactions. The mentioned high-level interactions are: selecting a Wikipedia article, reading a Wikipedia article, creating an annotation, cancelling to create an annotation, editing the annotation, cancelling to edit an annotation. The first two interactions (selecting and reading a Wikipedia article) are the interactions of gaining information. The other four interactions are annotation related interactions, where the user expresses his understanding from the obtained information. Thus, by visualizing these main 6 interactions, it is possible to see the flow from obtaining the information to showing user's understanding.

The *Figure 5* illustrates a high-level view of an annotation creation process for some paths. Each line corresponds to one path. The number at the start of each line shows the number of unique articles read in this path. The title of the article defined by the user is shown. This title probably indicates from which article the annotation is generated. After that, colored rectangles appear. Each rectangle represents reading an article, different colors refer to different articles in each path. The width of the rectangles

corresponds to reading time of the article. Hovering the rectangle, shows the corresponding Wikipedia title. The "x" symbol indicates that the user cancelled to create an annotation. The "Pen" icon with green tick implies that the user edited the annotation. The "Pen" icon with red cross indicates that the user cancelled to edit an annotation. Finally, the tick in the circle with green color represents, that the user created an annotation.
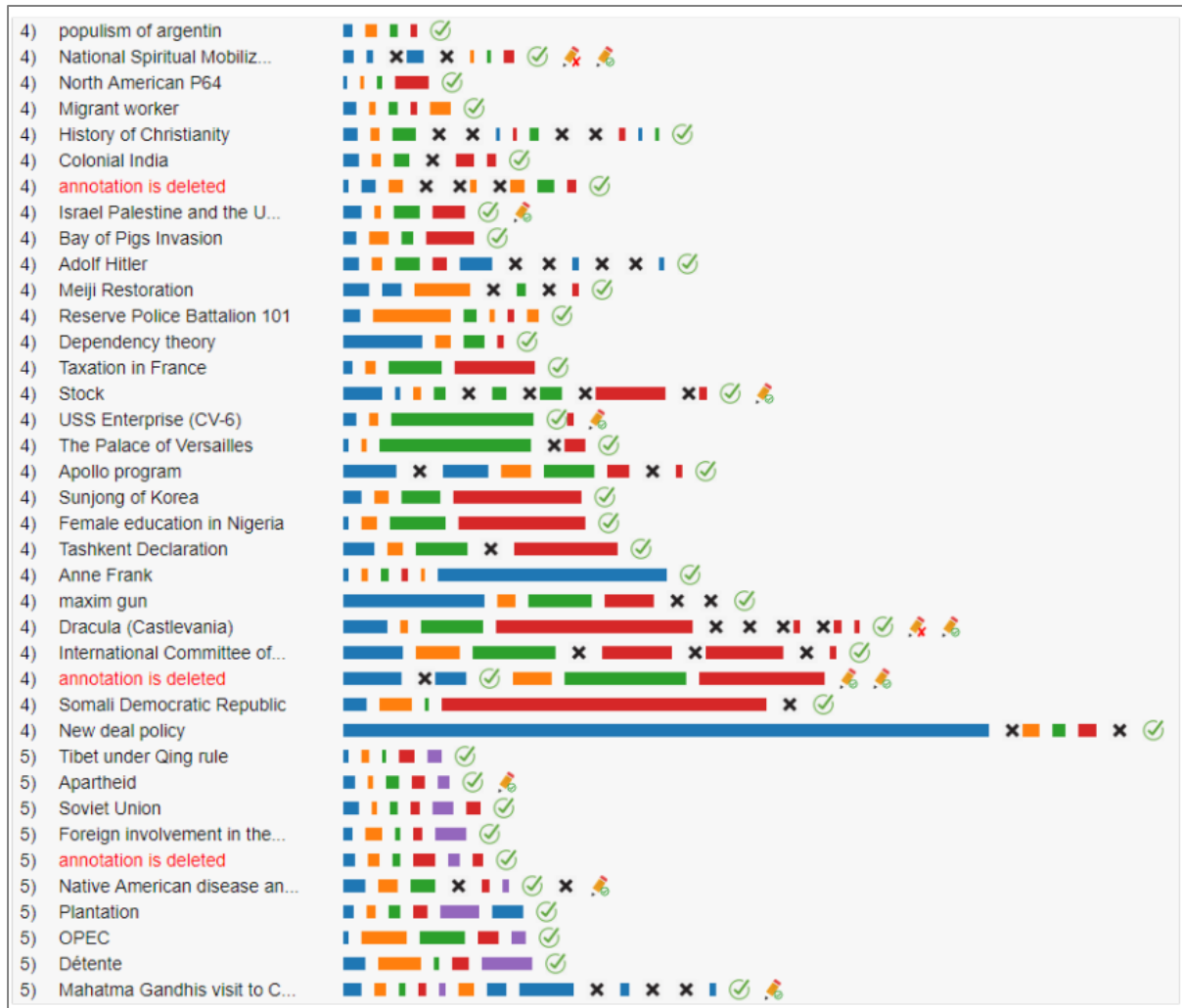

Figure 5: High-level view of an annotation creation process for some paths

This visualization shows that the annotation creation process is not just reading an article and creating an annotation. It shows that the number of unique articles is different for each path, starting from a single article to 85 articles. It also shows that the reading time is varied from few seconds (1.4 s) to more than an hour (73.6 min).

Although, the annotation creation process is different between paths, there are some common behaviors also. For example, most of the users cancelled creating an annotation several times (denoted by the "x" icons in Figure 5). Usually, they re-read the same article again to gain more information to finally create an annotation. The possible explanation for this behavior is that, after reading the article, the

users gained insight and started to formulate the insight in the mind. But, when they started to write the insight to the annotation form, they realized that some part of the information is missing. Therefore, the users cancelled creating an annotation, and re-reads the article.

Although the visualization (Figure 5) gives us detailed information about each path, it is difficult to see the overview of an annotation creation process. Therefore, this visualization is integrated into a graph in *Figure 6*, to see all the paths together.
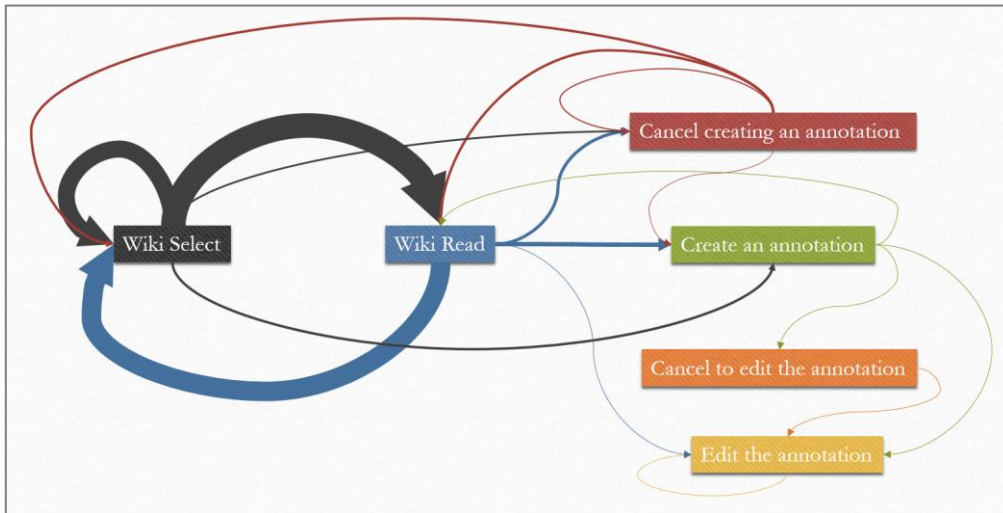

Figure 6: High-level graph of an annotation creation process

Figure 6 illustrates the iterative insight generation process, each node represents a Wikipedia article or annotation related interaction, the edge width corresponds to the frequency of a path from one interaction to another. It shows that the users mostly travelled between selecting a Wikipedia article and reading it. It supports the cyclic behavior of an insight generation process. The users read several Wikipedia articles one after another until they feel comfortable to make an annotation. But even, during the annotation creation process, the users cancelled to create an annotation and went back to read the same or another article. After several such kind of cycles, they finally created an annotation. However, sometimes, the users still edited the annotation, after creating it, without any extra article reading or reading some articles. It all supports the hypothesis, that the insight generation process is iterative.

To make Figure 6 more readable, I applied two filtering rules. First, we can consider the "wiki select" and "wiki read" interactions as one block "wiki", since these 2 interactions are for seeking information, and other 4 are for users to show their understandings. Second, we can remove some of the edges based on its existence among 546 paths because not all 546 paths contain all the edges in Figure 6. For instance, only 11 paths have the interaction sequence from "Cancel to edit the annotation" to "Edit the annotation", which is only 2% of total paths. These edges represent more individual behaviors rather than common features among the users. Therefore, I removed the edges with existence value of below

10%. The full list of existence values is given in *Table 2*. The result after applying the mentioned two filtering rules is shown in *Figure 7*.

First Interaction	Second Interaction	Existence in 546 Paths	Existence in Percent
wiki	created the annotation	502	92%
wiki	wiki	471	86%
wiki	cancel creating an annotation	206	38%
cancel creating an annotation	wiki	177	32%
cancel creating an annotation	cancel creating an annotation	68	12%
created the annotation	edit the annotation	52	10%
cancel creating an annotation	created the annotation	44	8%
created the annotation	wiki	19	3%
wiki	edit the annotation	19	3%
edit the annotation	edit the annotation	16	3%
created the annotation	cancel editing an annotation	12	2%
cancel editing an annotation	edit the annotation	11	2%
wiki	cancel editing an annotation	5	1%
cancel editing an annotation	wiki	5	1%
cancel creating an annotation	edit the annotation	3	1%
cancel editing an annotation	cancel editing an annotation	3	1%
created the annotation	cancel creating an annotation	2	0%
cancel editing an annotation	cancel creating an annotation	1	0%

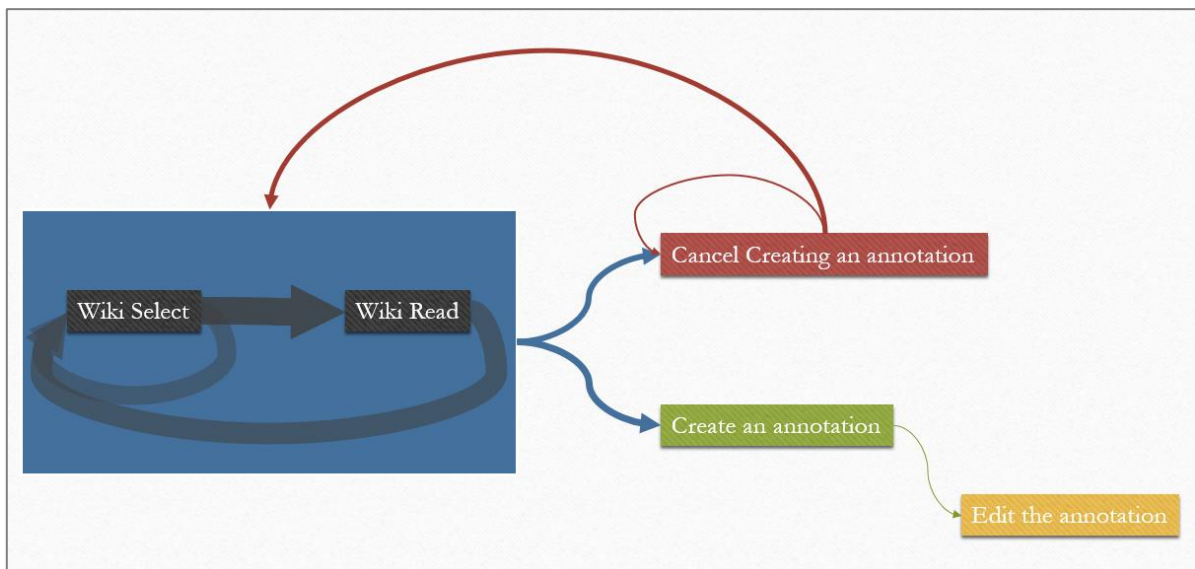Table 2: Existence of annotation and wiki related interaction pairs

Figure 7: Filtered version of a high-level graph of an annotation creation process

According to *Figure 7*, there are 2 main behaviors. The first one is creating an annotation directly after several cycles of selecting and reading the Wikipedia articles. Second, the users selected and read the Wikipedia articles several times, then cancelled to "create an annotation", then came back to read a new

10

or already visited Wikipedia article. After repeating this cycle several times, users finally created an annotation. These two behaviors are visualized as a separate diagram in *Figure 8*, with numerical values.
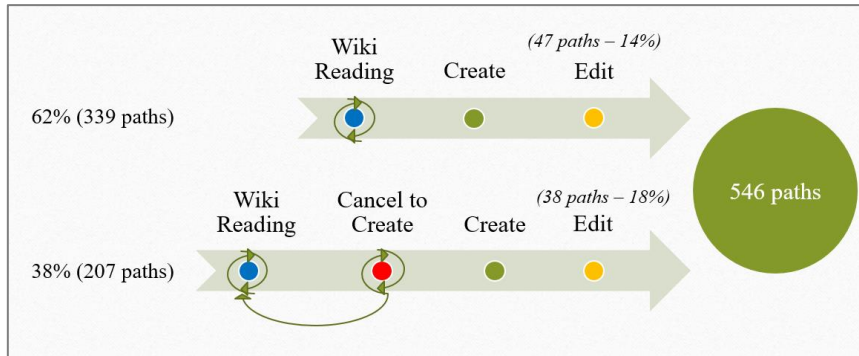

Figure 8: Two common behaviors in annotation creation process

The diagram shows that 38% of the paths include the "cancel to create" interaction. It implies that the cyclic behavior of insight generation process is not just a random or seldom case, it is a considerably common pattern among the users.

4.3 The Role of Visualization Interactions in Insight Generation Path

Although, the high-level view of an insight generation process gives an opportunity to observe the relationship between Wikipedia article reading and annotation creation, it is worth to explore the role of other interactions such as map view, timeline view in an insight generation path.

To visualize the common paths from 546 paths with all interactions is not a trivial work because some have too many interactions in a sequence. However, as we discovered from the high-level view of an insight generation path, the users showed a cyclic behavior in terms of Wikipedia article selection. Thus, if we pick the "wiki select" interaction as a pivot and visualize 3 different visualizations: 1) The interactions from path start to first Wikipedia article selection, 2) the interactions between Wikipedia article selections and 3) the interactions from last Wikipedia article selection to annotation creation, it gives an opportunity to visualize the graph efficiently. Furthermore, to reduce the number of nodes in the graph and keep only common interactions, I filtered all interactions based on two criteria. First, I removed some of the low-level interactions such as "scrolling" or "clicked annotation icon". Second, I only kept the interactions that exist at least in 55 paths (10%) among 546 paths. As a result, there are now only 20 unique interactions out of 138 initial interactions. The list of the filtered interactions is shown in *Table 3*.

Interaction name	The number of paths where this interaction exists
map0 view->created the annotation	539
map0 view - title table->selected	536
wiki1 view->elapsed time	468
map0 view->navigated	337
search box->article	336
timeline view->selected a time range	278
right tab->changed to AnnoGroup	270
topic tree view->navigated	255
login form->logged in	245
map0 view->cancel creating an annotation	227
topic tree view->selected topic	217
right tab->changed to Annotation	203
map0 view - location table->selected	183
right tab->changed to Keyword	169
map0 view->clicked blue W	121
topic tree view->searched	118
topic circular view->navigated	79
topic circular view->selected topic	67
timeline view - slider->changed time graph range	65
map0 view->opened wiki article	58

Table 3: The list of the filtered interactions
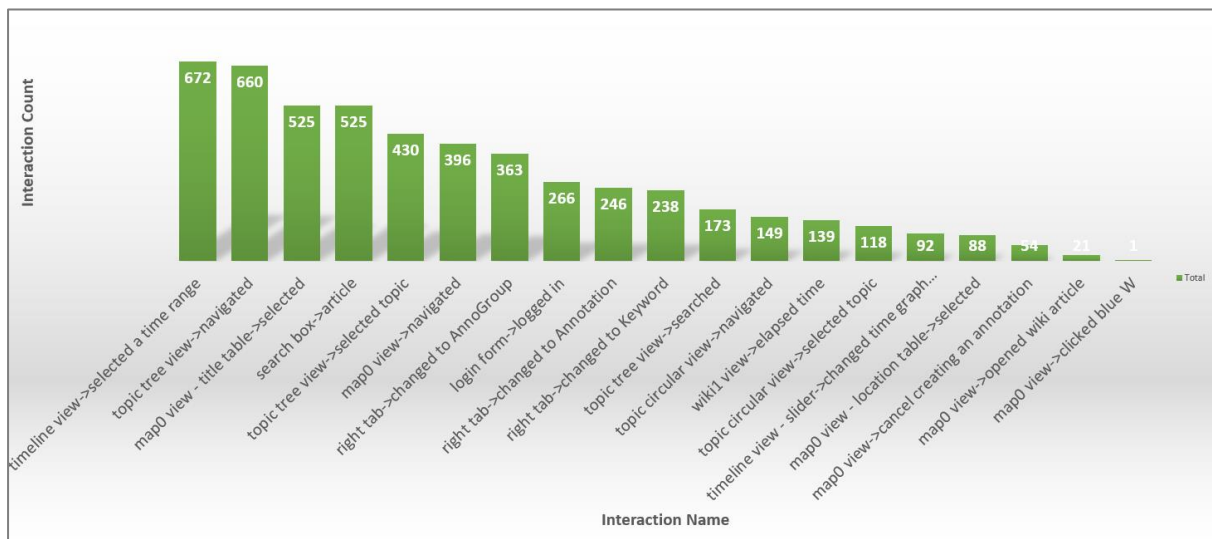
4.3.1 Path to First Wikipedia Article Selection

Figure 9: The total number of each interaction before selecting the first Wikipedia article

To discover how users found the first Wikipedia article, I started analyses by visualizing frequency of the interactions (*Figure 9*). The frequency of interaction indicates the total number of occurrence of each interaction before selecting the first Wikipedia article in all 539 paths. For example, from Figure 9, we

can see that "timeline view->selected a time range" occurs 672 times. Since this is the most frequent interaction, we may guess that the timeline view had a big role toward first Wikipedia article selection. Other frequent interactions are navigating in topic tree view, selecting Wikipedia article, searching Wikipedia article from a search box, selecting topics from topic tree view etc. All these frequent interactions are reasonably essential interactions in a path to find a Wikipedia article. However, this graph is a high-level observation and it may blur some important information. For example, there is a possibility, that all 525 "search box->article" interactions were performed in a single path, and other paths may not contain these interactions. To reveal this feature, I draw another graph in *Figure 10* which shows the existence of each interaction in all paths, before selecting the first Wikipedia article.
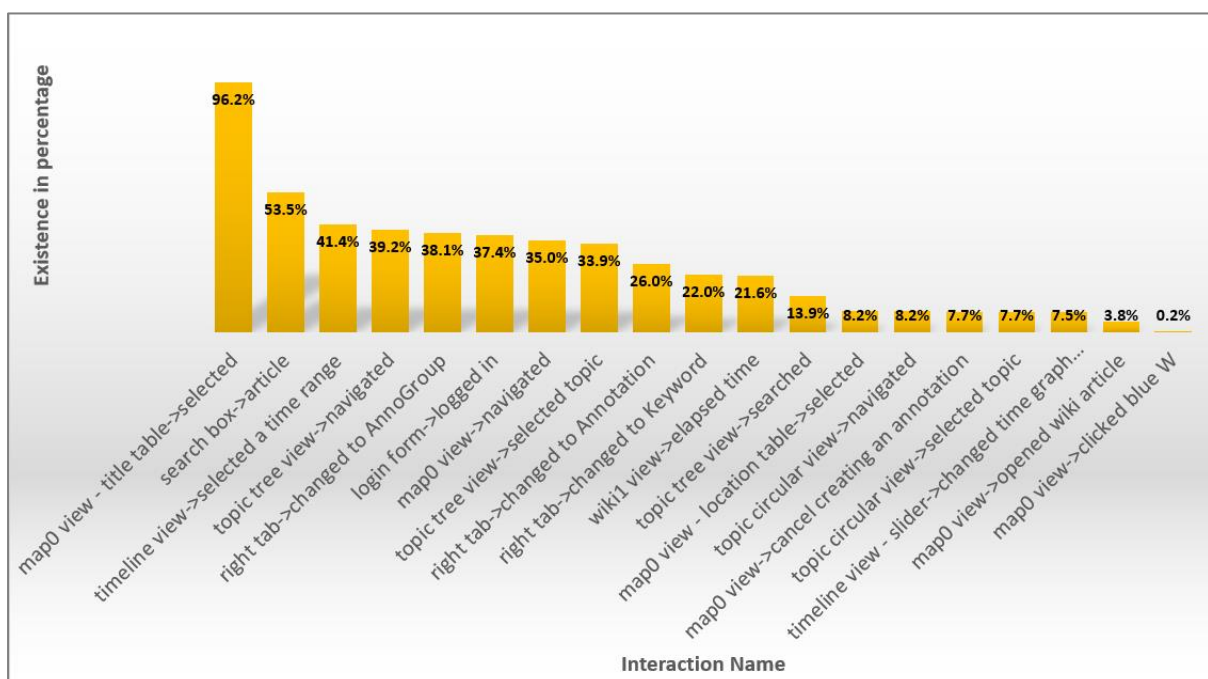

Figure 10: The existence of each interaction before selecting the first Wikipedia article

According to *Figure 10*, the "map0 view – title table->selected" interaction is 96.2%. Logically, this interaction need to be 100% since we are considering the interactions before the first wiki selection (including). But, other 3.8% interactions happened in different view, rather than "map0 view". Let's move to the second highest interaction "search box-article" (53.5%). It means searching the article by the search box was a major interaction to find the first Wikipedia article. The interaction "timeline view->selected a time range" also leads to a Wikipedia article list. However, the users may have preferred to use the search box than the timeline view for first Wikipedia article selection.

To find out the path to first Wikipedia article selection, a tree-based visualization technique is used in *Figure 11*. The visualization shows two main paths. In the first case, a search interaction is held before selecting a Wikipedia article, which is 243 paths or 44.5%. In the second case, the Wikipedia article is found by a timeline view, which is 133 paths or 24.4%.
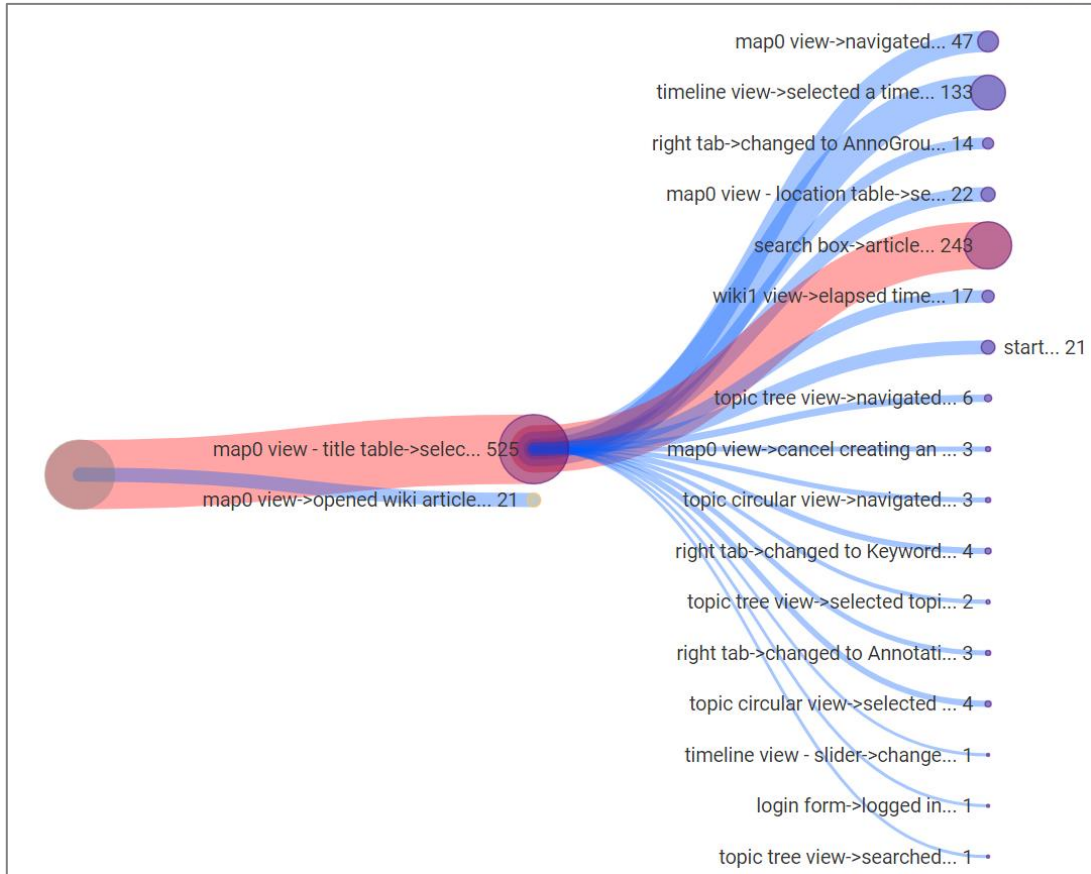
Figure 11: Interaction sequences from path start to select the first Wikipedia article

To observe the tree paths deeper, the expanded version is given in *Figure 13*. Here, I will only focus to the nodes with at least a value of 54, which is 10% of total paths. Because, this threshold guarantees us to extract a common pattern. After carefully analyzing the tree, I realized that there are 2 common paths, which lead to selecting the first Wikipedia article. First one is iteratively using the "search box" and another is navigating in topic tree view to study topics, then selecting a suitable topic, then using the time-line visualizations to obtain a list of Wikipedia articles. This behavior is drawn in *Figure 12*.
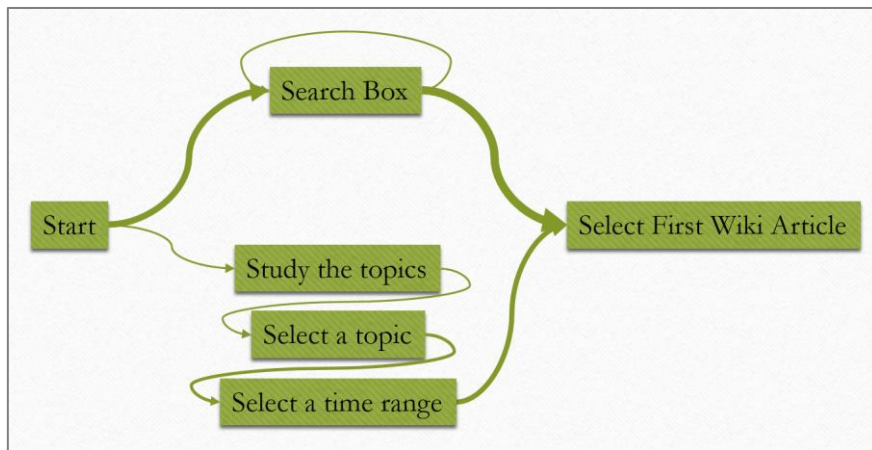

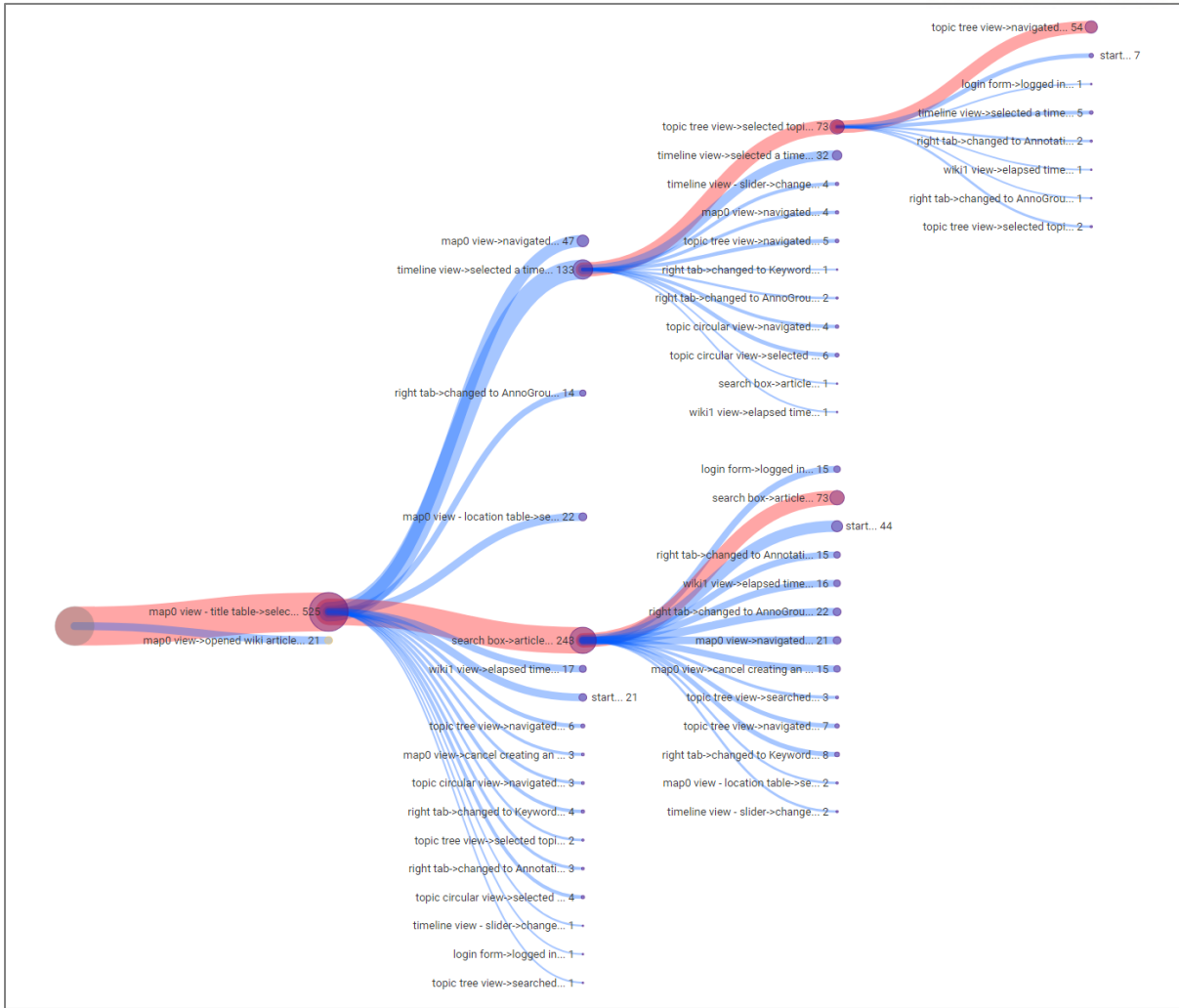Figure 12: The diagram of selecting the first Wikipedia article

Figure 13: The expanded version of tree visualization for finding the path to first wiki selection

4.3.2 Paths Between Wikipedia Article Selections

In this section, I will analyze the interaction sequences from one Wikipedia Article selection to another. By conducting these analyses, we may understand why users selected several Wikipedia articles and what were the main interactions between Wikipedia article selections.

I started the analysis by visualizing frequency of interactions in *Figure 14* and existence of the interactions in each path in Figure *15*. The most frequent interaction is selecting a Wikipedia article. Since we are considering the interaction sequences from first Wikipedia article selection to last Wikipedia article selection in a path, it is reasonable that selecting a Wikipedia article is the most frequent interaction. There are two interesting patterns. First, reading a Wikipedia article (wiki1 view->elapsed time) become one of the most frequent interactions. Second, the frequency of "search box->article" decreased considerably comparing to "topic tree view" or "timeline view" interactions. Thus, there is a possibility, that the search box is used frequently only for the first Wikipedia article selection, but when the users explored several Wikipedia articles, another way of article finding, which is topic
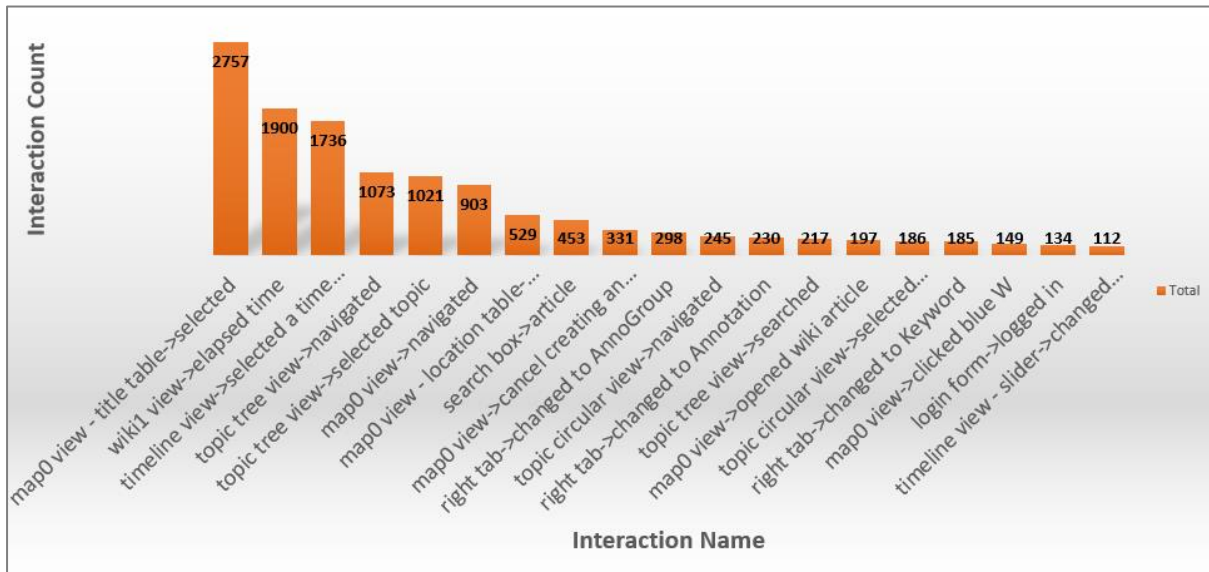
tree view and timeline view is used more.

Figure 14: The total number of each interaction before selecting the first Wikipedia article
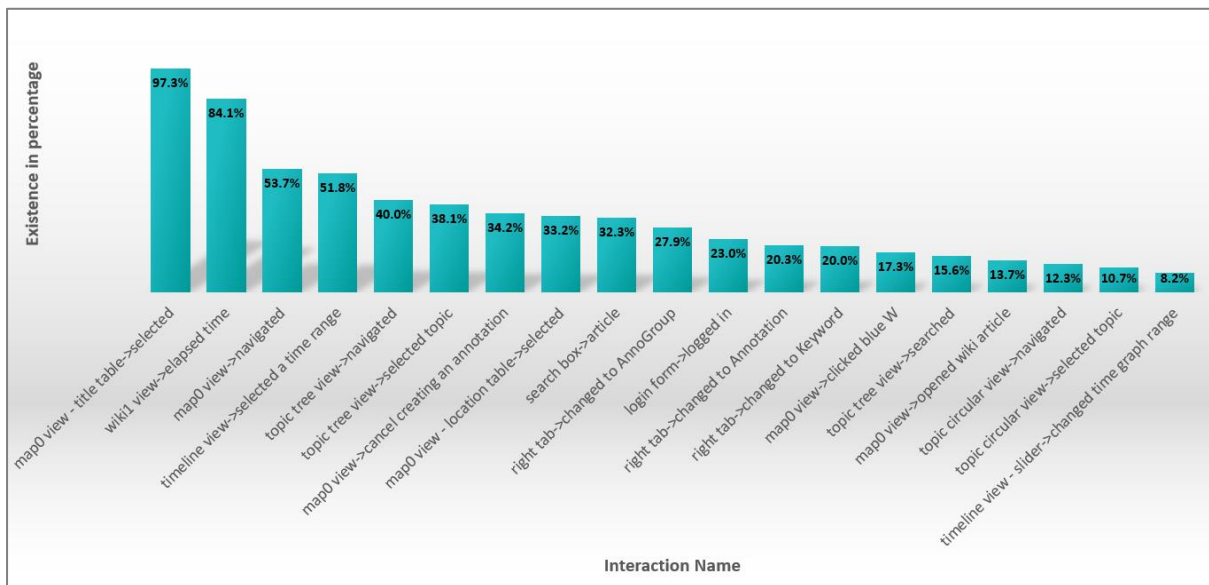

Figure 15: Existence of the interactions in sequences from first to last Wikipedia article selection

Based on the results of *Figure 14* and *Figure 15*, we can predict that when the users selected several Wikipedia articles, between the selections, usually they read the selected Wikipedia article, did explorative interactions such as navigating in map view, topic tree view, selecting a time range and selecting a topic. Also, they tried to create annotations, but cancelled to create. One interesting pattern is they did not use the "search box" frequently comparing to the first Wikipedia article selection. To see the common interaction sequences from one Wikipedia article selection to another, I visualized the path tree in *Figure 16*. According to this graph, there are three main paths between Wikipedia article selections. First, the users selected a Wikipedia article just after another selection. It indicates that the users selected

a Wikipedia article from already found article list. Second, the users moved their mouse over wiki view to read the Wikipedia article. Third, after reading the Wikipedia articles, in some cases, the users did a new exploration by the topic tree view, timeline view to find new Wikipedia articles. These three main paths are drawn as a diagram in *Figure 17*. Different colors refer to different paths.
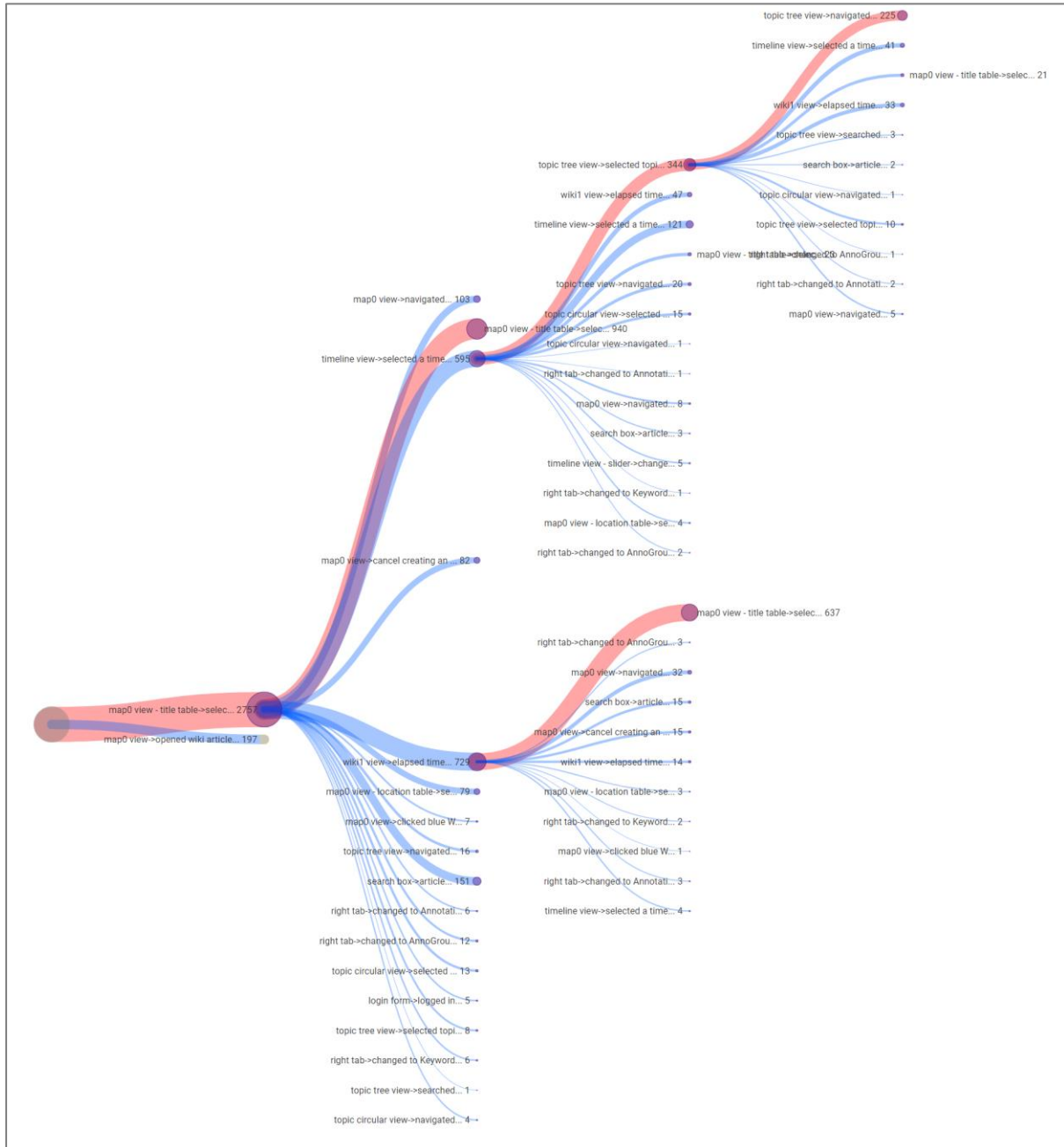

Figure 16: Tree visualization for finding the path from one wiki selection to another
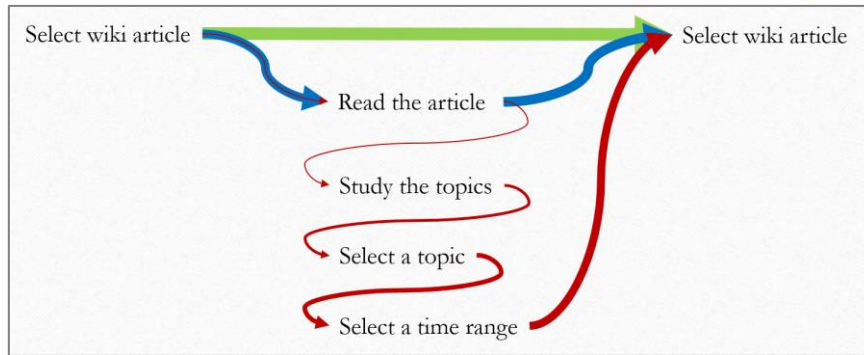
Figure 17: The diagram of 3 main paths between Wikipedia article selections

4.3.3 Path from Last Wikipedia Article Selection to Annotation Creation

In this section, I will analyze the interaction sequences from last Wikipedia article selection to annotation creation. This might be the most important part of an insight generation path due to the annotation creation interaction. According to *Figure 18*, between Wikipedia article selection and annotation creation the most frequent interaction was to move the mouse over the wiki view to read the Wikipedia article (wiki1 view->elapsed time). Also, navigating in map view was one of the most important interactions toward annotation creation. This can be supported with *Figure 19*, which shows the existence of each interaction in sequences from last Wikipedia article selection to annotation creation.
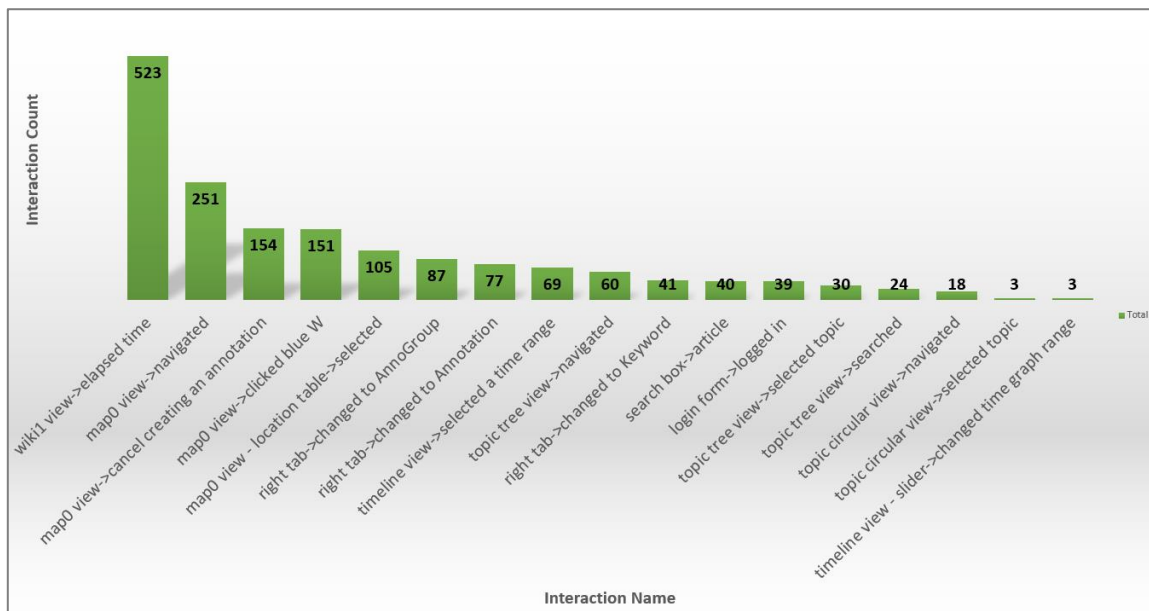

Figure 18: The total number of interactions from last wiki selection to annotation creation
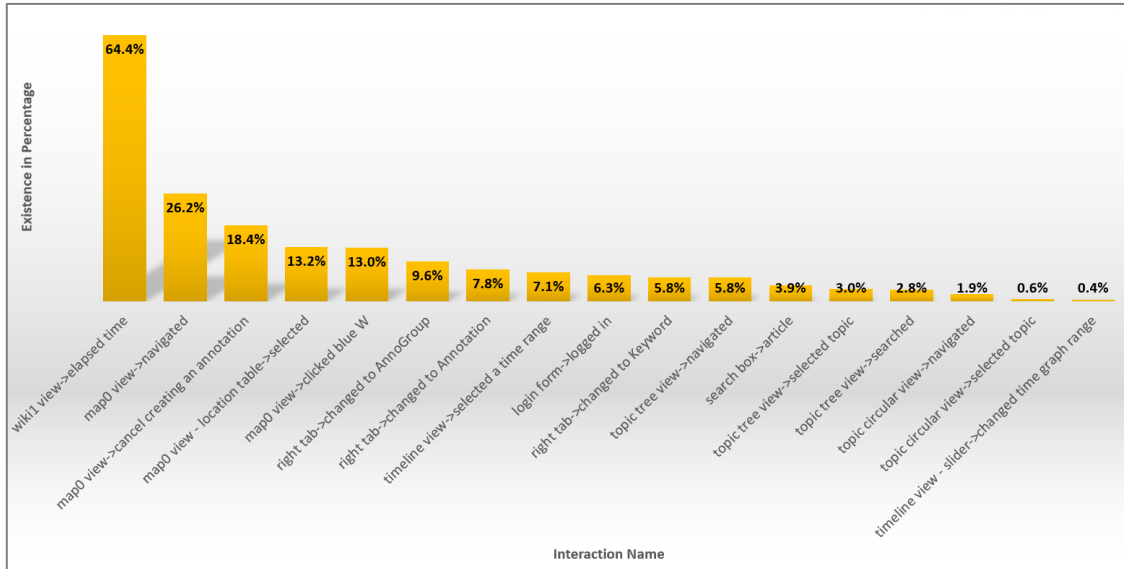
Figure 19: The existence of the interactions from last wiki selection to annotation creation
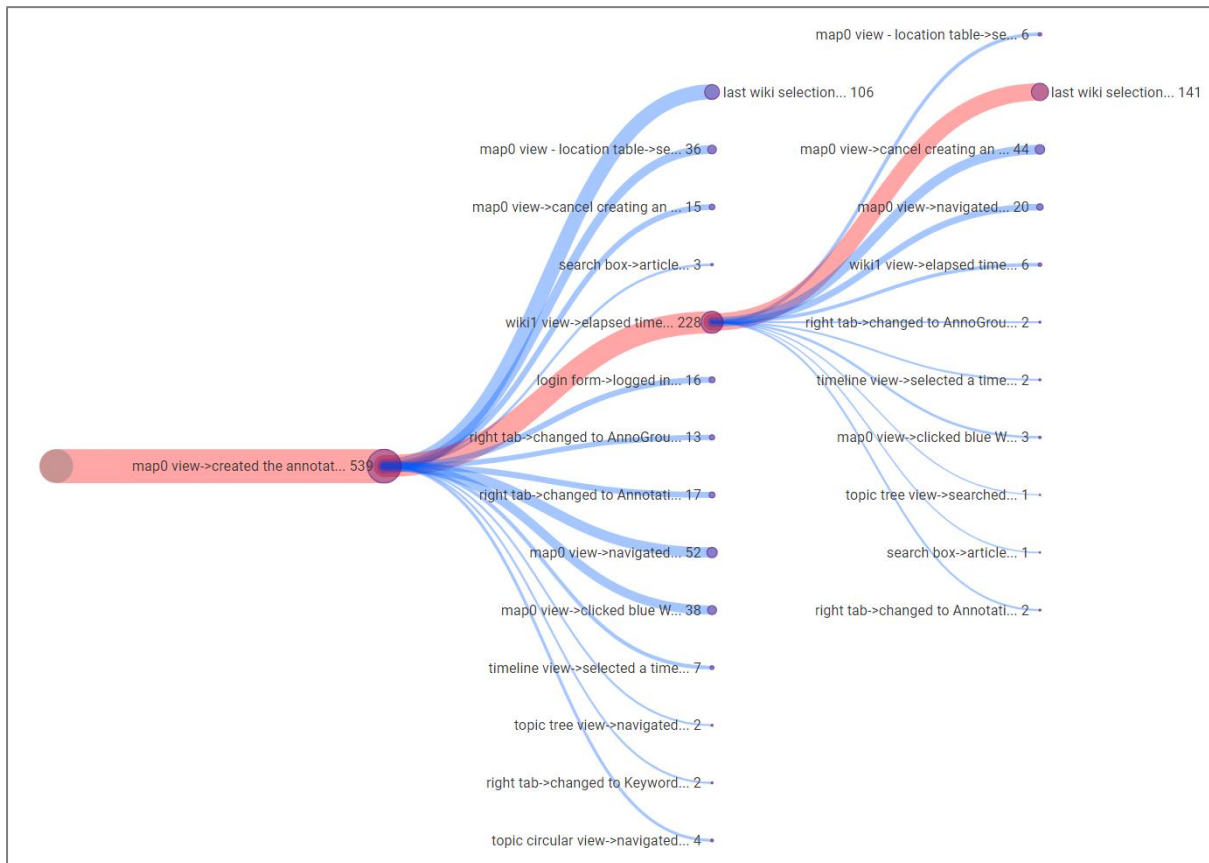
Figure 20: Tree visualization for finding the path from last wiki selection to annotation creation

The tree visualization in *Figure 20* shows the main path from last Wikipedia article selection to annotation creation. According to it, after selecting the last Wikipedia article, the users moved their mouse over wiki view (wiki1 view->elapsed time) to read the selected Wikipedia article and then created the annotation. Another major path was to create the annotation right after selecting the

Wikipedia article. One interesting pattern is the frequency of explorative interactions such as the topic tree view or timeline view or search box is almost negligible. It may imply that the users already finished all explorative interactions at initial stages and ready to create an annotation. To conclude this part a diagram of common paths is drawn in *Figure 21*.
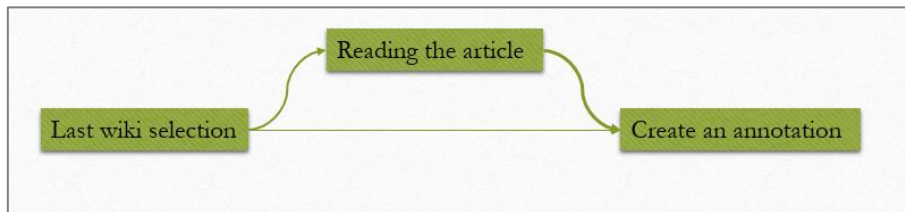

Figure 21: Diagram of common paths from last wiki selection to final annotation creation

V. Discussion

In the previous section, I analyzed the user interaction logs of an educational visualization system from two different aspects. First, I picked "Wikipedia article" and "annotation creation" related interactions to see the high-level view of an insight generation path. It gave us an opportunity to observe the effect of Wikipedia article selection and article reading to the annotation creation process. According to the results, we detected the iterative behavior of an insight generation path. In 38% cases, users cancelled annotation creation several times to read more Wikipedia articles before creating a final annotation. Second, I analyzed the effect of other visualization interactions to annotation creation process. We can integrate these two analyses to see the overall architecture of insight generation process (*Figure 22*) in our system. The interactions in red color are the explorative interactions, blue – Wikipedia article selection and reading, green – annotation related interactions. The edge width corresponds to the frequency of the paths.
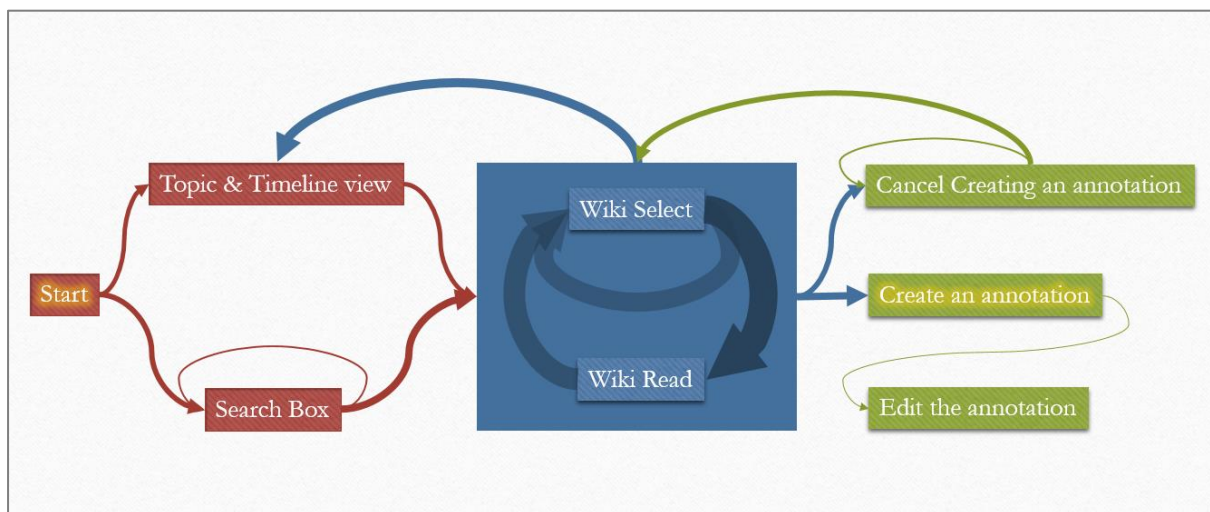

Figure 22: The overall diagram of an insight generation process

Users started the system with explorative interactions such as studying and selecting topics in topic tree view, selecting a period from timeline view or searching for Wikipedia articles directly by a search box. After that, the users moved toward selecting and reading Wikipedia articles. Sometimes, they came back to the explorative state again to find new Wikipedia articles. During the article reading state, they tried to create an annotation several times, but cancelled it. This might be due to the lack of information for annotation creation. After several such kind of iterative processes, they finally created the annotation. In 16% cases they edited the already created annotation. Based on these results, we may argue that the insight generation process is complex, and users always come back to previous state to gain more insights. Therefore, we should consider the cyclic behavior of an insight generation process when designing educational visualization systems. Examples could be providing bookmarks or history savings to make an easier access to already visited information.

Although I analyzed the user interaction logs from various aspects and obtained some results, there are still more to explore. One of the future works could be studying user interaction logs based on different cycles to observe user behavior change over time. Another interesting research might be to see a possible effect of different insight generation paths to the final annotation quality.

VI. Conclusion

In this thesis work, I tried to discover the insight generation path in educational visualization systems. To do so, I analyzed 42935 pre-processed user logs with 138 unique interactions. According to our results, the users explored the system in two different ways. First, using a search box to find a Wikipedia article. Second, using visualization tools such as topic tree view, timeline view and map view. In case of a search box, it was frequently used to find the first Wikipedia article or a single article. But, when users explored several Wikipedia articles, they preferred to use the visualization tools to access a wider range of article list.

One interesting pattern was, in 38% of cases, the users did not create the annotation directly after explorative state. They tried to create an annotation but cancelled to create and went back to find or read more articles. It implies the iterative behavior of an insight generation process. To fully understand something, it requires several cycles of exploration. Thus, this cyclic behavior of insight generation path should be considered when designing educational visualization systems.

REFERENCES

[1] I. Gutenko, K. Dmitriev, A. E. Kaufman and M. A. Barish, "AnaFe: Visual Analytics of Image-derived Temporal Features - Focusing on the Spleen," *IEEE Transactions on Visualization and Computer Graphics,* vol. 23, no. 1, pp. 171-180, 2017.

[2] X. Zhang, Y. Wang and P. Lv, "IBBAS: A Visual Analytics System of Large-Scale Traffic Data for Bus Body Advertising," in *International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, 2017.

[3] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu and Y. Wu, "SmartAdP: Visual Analytics of Large-scale Taxi," *IEEE Transactions on Visualization and Computer Graphics,* vol. 23, no. 1, pp. 1-10, 2017.

[4] E. S. Novikova, Y. A. Bekeneva and A. V. Shorov, "Towards visual analytics tasks for the security information and event management," in *International Conference "Quality Management,Transport and Information Security, Information Technologies"*, 2017.

[5] G. Silahtaroglu and H. Donertasli, "Analysis and prediction of E-customers' behavior by mining clickstream data," in *IEEE International Conference on Big Data*, 2015.

[6] I. Cho, W. Dou, D. X. Wang, E. Sauda and W. Ribarsky, "VAiRoma: A Visual Analytics System for Making Sense of Places, Times, and Events in Roman History," *IEEE Transactions on Visualization and Computer Graphics,* vol. 22, no. 1, pp. 210-219, 2016.

[7] P. H. Nguyen , K. Xu, A. Wheat, B. W. Wong, S. Attfield and B. Fields, "Sensepath: Understanding the Sensemaking Process through Analytic Provenance," *IEEE Transactions on Visualization and Computer Graphics,* vol. 22, no. 1, pp. 41-50, 2016.

[8] K. B. Chul, V. Janu and P. Adam, "Peekquence: Visual Analytics for Event Sequence Data," in *ACM SIGKDD Workshop on Interactive Data Exploration and Analytics*, 2016.

[9] F. Dabek and J. J. Caban, "A Grammar-based Approach for Modeling User Interactions and Generating Suggestions During the Data Exploration Process," *IEEE Transactions on Visualization and Computer Graphics,* vol. 23, no. 1, pp. 41-50, 2017.

[10] F. Du, C. Plaisant, N. Spring and B. Shneiderman, "EventAction: Visual Analytics for Temporal Event Sequence Recommendation," in *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, 2016.

[11] X. Chen, S. D. Starke, C. Baber and A. Howes, "A Cognitive Model of How People Make Decisions Through Interaction with Visual Displays," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 2017.

22

[12] H. Guo, S. R. Gomez, C. Ziemkiewicz and D. H. Laidlaw, "A Case Study Using Visualization Interaction Logs and Insight Metrics to Understand How Analysts Arrive at Insights," *IEEE Transactions on Visualization and Computer Graphics,* vol. 22, no. 1, pp. 51-60, 2016.

[13] M. Pohl, G. Wallner and S. Kriglstein, "Using Lag-Sequential Analysis for Understanding Interaction Sequences in Visualizations," *International Journal of Human-Computer Studies,* vol. 96, pp. 54-66, 2016.

[14] Y. Han, G. D. Abowd and J. Stasko, "Requirements for Visual Interaction Analysis Systems," in *IEEE VIS Workshop on Logging Interactive Visualizations & Visualizing Interaction Logs*, 2016.

[15] G. Wang, X. Zhang, S. Tang, H. Zheng and B. Y. Zhao, "Unsupervised Clickstream Clustering for User Behavior Analysis," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 2016.

[16] K. Reda, A. E. Johnson, M. E. Papka and J. Leigh, "Modeling and Evaluating User Behavior in Exploratory Visual Analysis," *Information Visualization,* vol. 15, no. 4, pp. 325-339, 2016.

[17] D. Gotz and M. X. Zhou, "Characterizing Users' Visual Analytic Activity for Insight Provenance," in *Proceedings of the IEEE Conference on Information Visualization*, 2009.

[18] Y. Chen, J. Yang and W. Ribarsky, "Toward Effective Insight Management in Visual Analytics Systems," in *Proceedings of the IEEE Symposium on Pacific Visualization*, 2009.

[19] J. S. Yi, Y. a. Kang, J. T. Stasko and J. A. Jacko, "Toward a Deeper Understanding of the Role of Interaction in Information Visualization," *IEEE Transactions on Visualization and Computer Graphics,* vol. 13, no. 6, pp. 1224-1231, 2007.

Acknowledgement