A Timeline News Algorithm-
A Better Way to Organize Online News

DongHoon Han
dhh96@khu.ac.kr
International Studies
KyungHee University

Abstract

There is a need for a better way to organize online news that helps people to see the whole picture of an issue in a single webpage. This study suggests a *Timeline news algorithm*, which connects particulated news into one whole story line by using the *linear storytelling structure (Eunyoung, Kang 2010)*. The algorithm divides the entire timeline of a fixed issue into multiple time intervals so that sub-issues can be grouped and organized in a linear fashion: First, extract related keywords of each date. Second, compare keywords with dates and group them as a sub-issue. Third, pick time intervals that are considered significant and identify the important stories. We test the algorithm with the keyword *Burning Sun*, a scandal that is very popular these days. As a result, we find that storytelling structure is helpful to understand the news.

1. Introduction

Advent and development of Internet technology greatly affected human society. Internet technology made production and consumption of data easier. This affected the whole environment of media including news media. Before wide spread of internet, only mass consumption of media was available through unidirectional transmission of signal such as radio or TV. However, producing and sharing of data has become easier as era of internet began.

Also, Internet technology enhanced dependency of news media. In the past, news media were distributed by producers. For example, broadcasting station creates its own news and broadcasts it. Nowadays however, distributors and producers of internet news media mostly do not match. News data are typically distributed by portal website except for some big and authorized publishing corporation (Wi-Geun Kim, 2014).

A portal, playing a role of news distributor these days, cuts expense for producing news since producers have no concern on distribution expense. Reduction in cost made news media gate-opening rather than gate-keeping (Kyungmo Kim, 2012). As a result, news creators increased and news media became exceedingly abundant.

Due to change in ecosystem as mentioned above, there are lots of news sources and contents on the internet nowadays. Most of news creators are dependent to portal websites (Yeo-kwang Yoon, 2104). Dependency on portal website of news creators created power imbalance between producers and distributors (Haeyeop Song, Jay Yang, 2017). From power imbalance, producers had to cater the distributors' need. On website, more click means more money. So, all those news producers tend to create short, quick news to compete with their competitors; to get more clicks from users and suit distributors' need. "Competition of click" led to creation of bunch of low-quality news contents focusing on temporary issues. Therefore, with online news it is becoming harder to see the whole picture of an event. For example, there are lots of news regarding "Burning sun" in South Korea. "Burning Sun" is a club in Gangnam, South Korea. "Burning Sun" started to get spotlight by assault incident. Burning sun issue got even more spotlight as more scandals like "Selling drug", "Sexual favors", "Hidden camera", and "Back-scratching relationship between politician and police" being revealed to public (Joan, MacDonald 2019). Burning Sun scandal caused so many sub-issues and it is actually impossible to follow what has happened from the start. Also, online news that focuses on temporary issues does not help much. In other words, it is hard to see the wood for the trees from online news media. This study proposes "timeline news algorithm" that is designed to alleviate the limitation of online news. There are examples with Burning Sun issue at section 3. With examples provided on section 3, readers would be able to check out how much this algorithm would help to understand the Burning Sun issue.

There is a need for news contents that tells overall flow of an event as news posted on online media has limitations on understanding the whole issue. News format that could ease limitation of online news are called "story telling news" (Ji-Yeon, Kim and Jae Young, Yun 2015) but there is no service that consistently provides storytelling news. To suit those needs, this study proposes "Timeline news algorithm" that connects particulated news into one whole story line to help people see the whole picture of an issue. "Timeline news algorithm" uses "linear storytelling" structure. Linear story telling structure (Eunyoung, Kang 2010) is a structure that tells what happened in time series. Since "Linear story telling" structure provides information in chronological order, it is very intuitive and simple. Linear story telling structure is easy to comprehend information on web. Algorithm using characteristics of linear story telling structure will cover the sub-issues from time to time about certain issue to help peoples' understanding. By searching sub-issues, algorithm will enable readers to

follow updating sub-issues (or minor scandals regarding issue) along with the whole picture of event.

2. Design of Timeline News Algorithm

To briefly show what timeline news algorithm does, following is expected result of timeline news algorithm. Timeline news algorithm will make multiple number of time lines. Each block has its own sub-issue, time interval, and articles. Algorithm will disassemble one big issue into multiple number of sub-issues. It will enable users to better understand about issue efficiently by breaking and gathering minor sub-issues.

For better understating about timeline news algorithm's structure, there is a hypothetical image of algorithm. Figure 1 is how this algorithm result should look like. There are two axes in the Figure 1 ((b) and (e)). Both axes are chronological timeline. Horizontal axis helps people to see the whole event (all the sub-issues) at once by showing overall period of an issue. Random dates are inserted in horizontal axis as an example. Vertical axis helps people to read specific articles which would be helpful for understanding details about sub-issues. Users can adjust (c) to choose the number of result timeline block.

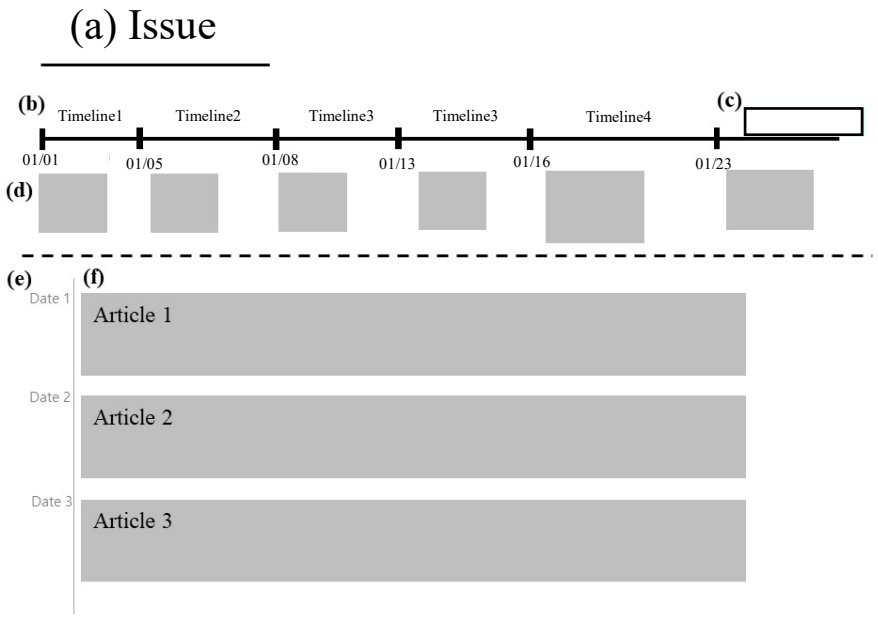

Figure 1. (a) issue of the timeline; (b) horizontal timeline; (c) is number of timelines blocks that user wants to see, (d) related picture of each timeline block; (e) vertical timeline; (f) space for news articles.

This chapter tells how algorithm works. 2.1 of the chapter works as main function of this algorithm. 2.1 contains all other processes. Figure 2 is the flowchart of 2.1 that tells general flow of this algorithm. 2.1 is composed of four big parts. First part is extracting related keywords of each date. 2.2 of this chapter will cover extracting related keywords of each date. Second part is comparing keywords with adjacent dates and expanding them into one time interval if they deals with same sub-issue. 2.3 and 2.4 of this chapter will cover comparing keywords and expanding time interval. Third part is prioritizing time interval blocks and erasing blocks that are considered less significant. 2.5 and 2.6 of this chapter will cover prioritizing process but this part will not be activated if user set rephrasing option "off". Last part is news searching process. 2.7 will cover this part.

2.1 Creating a Storyline of a News Issue

Getnews function works as main function for creating timeline news. The function returns timeline news data. Expected result of timeline news algorithm is each time interval having its own sub-issue (different with other interval's sub-issue). To be more specific, there will be multiple time intervals. Inside each time interval, there will be three related keywords that reflects time interval's sub-issue and news articles in each time interval. Timeline is divided as new sub-issue arises; dividing phase.

This function will return given number (option) of time intervals. If you set option as -1, function will return all the time intervals it makes. It was made to prevent the timeline from being long-winded. If you use this option, function will return time intervals that were spotlighted most by comparing each time interval's weight sum. This rephrasing option would enable users to see whole issue briefly but there would be some missing sub-issues since time intervals with less weight is removed.

To follow the flow of function, refer Figure 2 and Table 1. After **keywordextract** function in Figure 2, algorithm will return list of elements and each element will have date and relevant keyword of "Burning Sun". Of course, related keywords will gradually change as new sub-issues arise.

After **Phasedivide** algorithm, all the dates will be combined to multiple numbers of time intervals. Intervals are made if adjacent dates seem to have same sub-issue. Since dates in the interval deal with common issue within a time interval, it is likely that dates in same interval have common related keyword. To handle multiple number of common related keywords, **Keyword_Integrate** function put same related keyword together within time interval (weights will be added).

If rephrasing option is set to -1(off), then algorithm will jump to **NewsSearch** function using flow 1 in Figure 1. **NewsSearch** will search news for every time interval. When searching news, time period, issue, and three related keywords of the interval will be given. By adding three related keywords for search option, search result will show relevant news contents that contains those related keywords.

If there is a rephrasing option (positive integer N) algorithm will go through flow 2. Algorithm will add up all the related keywords' weight in each time interval and compare the value of weight sum of each interval. N time intervals with highest weight sum will be the input of **NewsSearch** function. Since Figure 1 is desirable outcome of this algorithm, Figure 2 was designed to return data for Figure 1. Overall result will carry period of each time interval, related keywords, and articles that could fit into Figure 1.

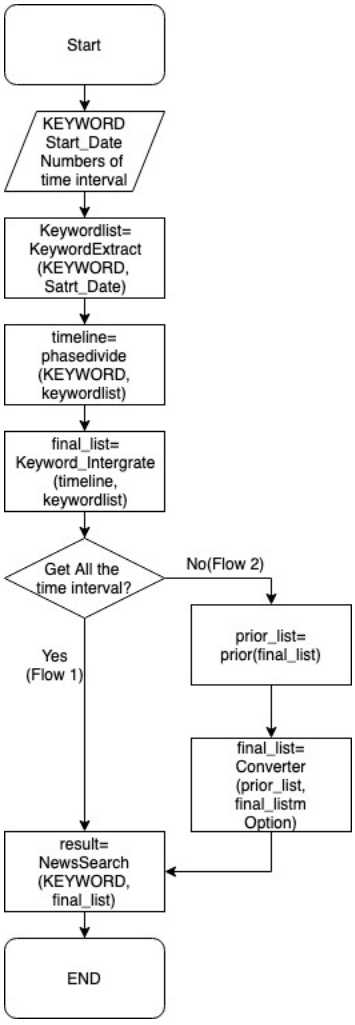

Figure 2. Flowchart of timeline news algorithm.

Table 1. Pseudo code that returns timeline intervals with articles and related keywords in it.

GetNews function

```
Input: Issue, BEGIN_SEARCHING, OPTION
keywordlist ← keywordextract(KEYWORD,BEGIN_SEARCHING)
timeline ← phasedivide(KEYWORD, keywordlist)
final_list←keyword_Integrate(timeline, keywordlist)
if (rephrasing option is off)
    prior_list←prior(final_list)
    input_list←converter(prior_list,final_list, option)
    article←NewsSearch(KEYWORD, input_list)
else
    article←NewsSearch(KEYWORD, final_list)
return article
```

2.2 Extracting Related Keywords of Date

This is the function that collects related daily keywords and its weights. This function used “timeline API” and “wordcloud API”. “Timeline API” checks number of news posted

on a designated day with given issue like “Burning Sun”. “Wordcloud API” uses both “Term Frequency” algorithm and “Inverse Document Frequency” algorithm to get related keywords and its weight.

By using “timeline API”, date with less than 10 posted regarding issue was not included in keyword extracting process (Considering that date has less or no relevance with given issue). Also getting rid of dates like this would make algorithm to finish timeline if case is closed.

At Table 2, **detect function** will return dates in list format. If less than 10 news were posted with issue, that date will not be included in the list. This function uses “timeline API” in Table 3. At Table 2, **getwordcloud function** collects related keywords and its weight. This function uses “wordcloud API” in the Table 4.

Table 2. Pseudo code that extracts relate keywords on daily basis.

Keywordextract function

```
    Input: Issue, BEGIN_SEARCHING
    Output: List of “date and that date’s related keywords”
1: labels ← detect(Issue, BEGIN_SEARCHING)
2: for all elements in labels:
3:     date ← get date information from element
4:     result ← getwordcloud(Issue, date)
5:     append result to final
6: return final
```

For better understanding, artificial example will be given below.
Suppose Ketwordextract function recognized {2019-03-03, 2019-03-04} as relevant date with input issue. Then, [[2019-03-03, {hotdog: 32.33, pizza:121}], [2019-03-04, {tomato: 10, rice: 65.1}]] will be given as result. In this example, related keywords and it weights are randomly made. Result of this function will be the input of 2.3 and return result will be saved in variable keywordlist in Figure 2.

2.2.1 Obtaining Relevant Dates with Keyword

Detect function at Table 3 uses “timeline API”. The variable timeline in line 2 of table 3 will have list object with dates and number of news posted as a return value of API request. This function is used to detect relevant dates with given issue.

Table 3. Pseudo code detects number of news posted with given issue.

Detect function

```
    Input: Issue, BEGIN_SEARCHING
1: Get today’s date
2: timeline ← timeline API request(Issue, from BEGIN_SEARCHING until today)
3: for all elements timeline”
4:     if (news posted >= 10):
5:         append element to result
6: return result
```

For better understanding, artificial example is given bellow. Suppose example is given with random date with following; [2019-01-01, 2019-01-02, 2019-01-04, ...2019-05-01]. Then this means that algorithm considers those dates were days that news related to issue were posted.

2.2.2 Getting Related Keyword with wordcloud API

Getwordcloud function uses “wordcloud API”. This function will return a date, related keyword of that date, and its weight. Function is designed to request daily wordcloud in this algorithm.

Table 4. Pseudo code that gets related keywords and its weight by using “wordcloud API”.

Getwordcloud function
Input: Issue, date 1: wordcloud ← **wordcloud API request**(Issue, date) 2: wordcloud ← sort related keywords in descending order with weights 3: if(more that 10 keywords): 4: remove rest of keywords 5: return wordcloud

For better understanding, artificial example is given bellow.
Suppose output example is given with random date and random word; [2019-01-01, {angry tomato: 23, blue paper: 4}]. It means that “angry tomato” and “blue paper” were related keywords with given issue on 2019-01-01. Function returns related keywords and its weight of single date.

2.3 Expanding Single Dates into Multiple Number of Time Intervals

Phasedivide function iterates through each date and compare related keywords with adjacent dates’ related keywords. If related keywords are similar, function considers those dates deal with same sub-issue and combine them into one time interval. At line 3 in Table 5, there is a condition for checking similarity. If list of related keywords of date “n-1” contains top three related keywords of date “n”, algorithm considers that those dates deals with same sub-issue. If sub-issue has not changed for one week, function intentionally divides time interval to see more details about that sub-issue. Purpose of intentional division of time interval is to prevent time interval from being long-winded.

Goal of this function is to divide whole events into minor stages (or sub-issues). Phasedivide function will enable users to see the whole event in several significant time intervals that carry sub-issues. Making time intervals will enhance understanding of whole event. To give an example with “Burning Sun”, this algorithm should create new time interval if main sub-issue changed from “hidden camera” to “drug selling” as new scandal arises.

Table 5. Pseudo code that expands dates into time intervals by comparing related keywords of adjacent dates.

Phasedivide function

```
    Input: keywordlist
    Output: list of starting dates of each time interval
 1: initialize counter to 7
 2: for all elements in keywordlist:
 3:     if (one of  8t h date's top 3 keyword is in the i-1 the date's keywordlist):
 4:         reduce counter
 5:         if (counter is zero):
 6:             append element's date into resultlist
 7:     else:
 8:         append element's date into resultlist
 9:         reset counter
10: return resultlist
```

For better understanding, artificial output example is given bellow.
Suppose output example is given with following; [2019-03-01, 2019-03-05, 2019-03-12]. This means that first time interval is from 2019-03-01 to 2019-03-04. Second time interval is from 2019-03-05 to 2019-03-11. Each date represents the starting date of each time interval. Return value of this function will be input value of 2.4's Keyword_Integrate function.

2.4 Integrating Daily Related Keywords into Period Related Keywords

Timeline is return value of phasedivide function (previous function). Goal of this function is to integrate daily dates and keywords into time interval unit keywords and dates. Therefore, return value will be a list of periods (not date) and that period's related keywordlist. In this process, list of multiple dates" related keywordlists will be integrated and ordered in descending order of weight again.

For better understanding, artificial example will be given bellow. Suppose there is a time interval from 2019-01-01 to 2019-01-03. Let 2019-01-01's related keywords were {"a":132.1, "b":33}. Let 2019-01-02's related keywords were {"b":100, "c":7.3}. Let 2019-01-03's related keywords were {"b":150.4, "a":14.1}. Then, this time interval will be 2019-01-01 to 2019-01-03 related keywords of {"b":283.4, "a":146.2, "b":133}.

Table 6. Pseudo code that handles same keywords within a timeline by adding its weights.
Keyword_Integrate function

```
    Input: timeline, keywordlist
 1: for i in range(1, len(timeline)):
 2:     start ← timeline[i-1]
 3:     end ← timeline[i]
 4:     weight←{}
 5:     for j in range(start+1, end, 2):          #temp will be the daily keywords
 6:         temp←keywordlist[j]                    #and weights between start and end.
 7:         for k in range(0, len(temp)):          # iterates through daily keywords
 8:             if (temp[k] in list):
 9:                 add weight(value) to pre-existing keyword key
10:             else:
11:                 add new keyword(key)
12:     sort key values of weight in descending order of weight
```

```
13:       append its weight to result
14: return result
```

2.5 Ranking Period Blocks with Sum of Weight Value

Prior function will operate if option input were given. Prior function is a function of flow 2 in Figure 2. Each time interval's weights of keyword are added in this function temporarily. Then function will compare the sum value of multiple number of intervals of weight and rank time intervals. Weight is a value that reflects how much a certain word was referred with given Issue. So, related keyword with higher weight can be seen as frequently mentioned sub-issue. In other word, weight could be interpreted as peoples' interest about certain keyword. By comparing sum of these value, function will identify which time interval's keyword had more attention. By comparing each intervals weight sum, algorithm will prioritize the multiple time intervals considering higher weight sum dragged more attention to public.

If there is OPTION value of positive integer (rephrasing function), algorithm will go through this function to check which time interval to return. Time interval block with relatively low sum value would be excluded at return value.

This stage will be executed through Prior function and Converter function at Table 7 and Table 8. Prior function will add the weight values and Converter function will delete the time intervals that are not considered significant. Return value of Keyword_Integrate function becomes the input value of Table 7's function.

Table 7. Pseudo code that adds all the weights of related keyword within each time interval.

Prior function

```
    Input: final_list
 1: prior←[None]* len(final_list)
 2: for i in range(0, len(final_list)):
 3:       if(i←even):
 4:            prior[i] ← final_list[i]      #copy period value
 5:       else:
 6:            prioir[i] ← sum of weights in final_list[i]
 7: result ←copy "prior"
 8: for i in range(0, len(result_list))
 9:       set first(1st value) and last(len(result_list)th value) to 0
10:       rank sum of weight and replace sum with its ranking
11: return result
```

Return value of Prior function will be the input of Converter function in 2.6 only when option value was given.

2.6 Removing Time Intervals That Are Not Significant Enough

Prior_list is return value of prior function at Table 7. This function is next function of Prior function which means that it is on flow 2 of Figure 2. Like mentioned above, this function deletes the time intervals if weight sum of time interval is relatively small compared to other time intervals.

Table 8. Pseudo code selecting few timelines that has biggest weight sum.

Converter function

```
    Input: prior_list, final_list, option
 1: iterator ← len(final_option)
 2: iterator ← change to option if option is smaller than len(final_option)
 3: result_list←[None]*(2*iterator)            #initilalize list with enough size
 4: save first two element of final_list to first two space of result_list
 5: save last two element of final_list to last two space of result_list
 6: index←3
 7: for i in range (3, len(final_list)-2, 2):
 8:       if(prior_list[i] < number-2)
 9:            result_list[index] ← final_list[i]
10:            result_list[index-1] ← final_list[i-1]
11:            index+=2
12: return result_list
```

2.7 Searching News with Given Period and Keywords

NewsSearch function uses “Search API” in the line 9 of Table 7. Purpose of this function is to search news with each time interval. NewsSearch function will search news with given period, issue, and three relate keywords with big weight. Then, news that contains those keywords will be found.

Table 9. Pseudo code that searches news with issue and three related keywords.

NewsSearch function

```
    Input: Issue, final_list
1: for all element in final_list:
2:      if (element is date):
3:           start ← get start date from element
4:           end ← get end date from element
5:      else:
6:           word ← [Issue, element’s top 3 keywords]
7:           article ← search(start, end, words)
8:           append article to result
9 : return result
```

2.7.1. Searching News with Searching API

Search news function used “Search API” to search for news with given period and words. In timeline news algorithm, API request is sent with period, issue, other three keywords. Then API is set to return two most related news. When doing test case, all other information like publisher and name of reporter was excluded since the core purpose of this algorithm is to help understanding what happened all along (contents of news were what mattered the most when understanding a topic).

3. Case Study

There are two output examples with “Burning Sun” issue in this chapter. First one is making timeline without rephrasing. According to Figure 2, first example goes through flow 1 in the flowchart. Second one is making timeline with 10 time intervals. According to Figure 2, second example goes through flow 2 in the flowchart. Return value of timeline news algorithm should have multiple time intervals and each time interval has three related keywords and two news articles. For following reason, article was skipped in this study. The purpose of test case is to check if this algorithm reorganizes changing sub-issues well for the understanding of an issue. Article is the simple result of news search using issue and related keywords. So, to check the effectiveness of this algorithm, it is more important to check the related keywords of each time interval. To boost readers’ understanding of test case result however, title of each time interval was included in Table 10 and 11.

To help verify effectiveness of algorithm, there are titles of articles (result of this algorithm) at the back. From this, reader would be able to check if this algorithm helps to understand the whole story line without spending much time.

3.1. Result of algorithm by using flow 1 at Figure 2 with Burning Sun issue.

Since first example uses flow 1, all of time intervals were returned as result. Table 10 is the result of algorithm with Burning Sun issue using flow 1 in Figure 2. According to Table 10, there were 35 time intervals in total until the testing day. This means that “Burning Sun” can be divided into 35 meaningful intervals since each interval tend to carry one representative sub-issue.

Table 10. Result of timeline news algorithm without rephrasing option (Start date: 2019-01-01, Keyword: Burning Sun).

Time interval	Related Keywords	Title of articles
1	폭행 사건 (Assault case) 성추행 (Sexual harassment) 성폭행 (Sexual assault)	1. 버닝썬 반박 "폭행 사건 인정..마약과 성추행 성폭행은 아냐" (“Burning Sun” reputes “Admits assault case but not sexual harassment and sexual assault”) 2. ‘승리 클럽’ 버닝썬, 디스패치 단톡방 보도에 "물뽕·성추행? 절대 동의 못해" (“Burning Sun”; belongs to Seung Ri, never agrees to usage of “GHB” and sexual harassment)
2	성관계 동영상 (Sex tape) 성폭행 (Sexual assault) 이문호 대표 (CEO Lee)[1]	1. ‘성관계동영상 촬영 인정’ (Admits illegal sex shooting) 2. 버닝썬 성폭행 논란, 이문호 대표 SNS 에 "안심하고 오셔도 됩니다" 글 올려(“Burning Sun”, Dispute on sexual assault, CEO Lee once said the “Burning Sun” is very safe.)

[1] CEO at club “Burning Sun”

3	성관계 동영상 (Sex tape) VIP룸 (VIP room) 참고인 신분(Witness)	1. 버닝썬 성관계 동영상` 클럽 직원 참고인 소환(“Burning Sun” sex tape, club’s staff summoned as witness) 2. ‘버닝썬 성관계 동영상’ 클럽 직원 참고인 소환조사(“Burning Sun” sex tape, club’s staff summoned as witness for investigation)
4	마약 투약 (Drug administration) 애나(Anna) 유착 의혹 (Suspicion of collusion)	1. ‘버닝썬’ 애나, 엑스터시 투약 적발…… 추방 결정 ‘불복’ (“Burning Sun” Anna, injection of ecstasy…… objection to expulsion) 2. 버닝썬 결국 계약 해지...클럽 MD '애나' 엑스터시 투약 의혹 (“Burning Sun” contract terminated…Anna(club MD) suspicion of ecstasy administration)
5	경찰관 (Police officer) 유착 의혹 (Suspicion of collusion) 압수수색 (Search and seizure)	1. 경찰 `버닝썬 유착 의혹` 은행계좌·통신 압수수색 (“Burning Sun” alleged collusion with police force. Search for seizure on bank accounts and communication record) 2. `버닝썬 유착` 계좌·통신 압수수색…… 현직 경찰관 여럿 포함돼 (“Burning Sun collusion” search for seizure on bank account and communication…… had several police officers in service)
6	성접대 (Sexual favor) 이문호 대표 (CEO Lee) 사내이사 (In-house director)	1. “버닝썬 이문호 대표 모발서 마약 양성반응 검출”..경찰, 승리 내사 착수 (Positive reaction on drug on CEO Lee’s hair. Police launch an investigation) 2. ‘승리 카톡’ 논란에 승리 “자진출두해 마약검사 받겠다”…경찰, 버닝썬 이문호 출국금지 (Seung Ri “hopes to voluntarily appear and get drug test”, CEO Lee forbidden to leave country)
7	성범죄 (Sexual offense) 해시태그 (Hash tag) 물뽕 (GHB)[2]	1. 비에이 영균, 승리 겨냥? 스태프 앞에서 모욕적 언사 “이 정도 해도 안되면…”(Celebrity YoungGyun reveals about SeunRi’s. Personality controversy upon SeungRi) 2. 비에이(Be.A) 영균, 승리 저격?… “비수 꽂더니 몇 배로 돌아가”(Celebrity Be.A posted “What goes around comes around…”)

[2] GHB(Gamma- Hydroxybutyric acid) is drug that cause hallucination.

8	이문호 대표(CEO Lee) 유착 의혹 (Suspicion of collusion) 마약 투약 (Drug administration)	1. '버닝썬' 이문호 대표, 10시간 경찰조사 뒤 새벽 귀가..."수사 협조했다" ("Burning Sun" CEO Lee got a police investigation for 10 hours and left at dawn) 2. 버닝썬 이문호 대표, 10시간 경찰 조사받고 귀가 ("Burning Sun" CEO Lee got 10 hours of police investigation and left)
9	유착 의혹(Suspicion of collusion) 마약 투약 (Drug administration) 탈세 의혹 (Suspicion of tax evasion)	1. 경찰, 버닝썬 탈세 의혹도 수사…"1 년치 장부 분석"(Police investigates on "Burning Sun" tax evasion) 2. 버닝썬 탈세 의혹도 수사 착수…"1 년 치 장부 분석 중"(Police investigates on "Burning Sun" tax evasion. Analyzing 1 year of financial record)
10	성접대(Sexual favor) 성범죄(Sexual crime) 각종 의혹 (Various suspicions)	1. 성접대와 탈세, 산더미 의혹에도 군대 간다는 승리…들끓는 여론(SeungRi announce to go to military despite all the controversies around him) 2. '의경 지원' 논란 승리 "25 일 입대할 것"…결국 의경도 불합격(SeungRi applied for police to serve military service but turns out that he failed)
11	이승현(Seung Hyun Lee)[3] 아레나 (Arena)[4] 성접대 논란 (Sexual favor)	1.'승리 성접대 의혹' 클럽 아레나 압수수색…승리, 25 일 육군 입대 가능할까?(SeungRi suspicion for providing sexual favor. Could he go to military on upcoming 25th?) 2. 승리, '성매매 알선' 혐의 입건(SeungRi Summoned for suspicion of arranging sex favor appointment)
12	정준영 (Jeong Jun-young)[5] 유착 의혹 (Suspicion of collusion) 경찰총장 (Police chief)	1. 민갑룡 경찰청장 "버닝썬 유착 의혹? 지위고하 막론 발본색원" (Kap-ryong Min a chief of police hopes eradicate all the suspicious police members related with "Burning Sun") 2. 승리 카톡방 '경찰총장' 거론된 총경 "정준영 모른다" 참고인 조사 (Chief officer "A"

[3] Korean Celebrity. Also known as Seung-Ri. He established club "Burning Sun".

[4] Club in Seoul. Also got spotlight as "Burning Sun" scandal arises.

[5] Arrested for "Hidden camera" and sharing videos through messenger. Seung-Ri was in that chat room too.

		mentioned on Seung Ri's messenger. Investigation on potential witness)
13	현직 경찰관 (Policeman in service) 피의자 (Suspect) 직무유기 혐의 (Suspicion on dereliction of duty)	1. '버닝썬 미성년자 출입사건' 담당 경찰 피의자로 입건(Minors entered club "Burning Sun". Policeman charged for backscratching) 2. '버닝썬 미성년자 출입사건' 담당 경찰, 피의자로…현직 첫 입건(Minor entering club "Burning Sun". Policeman charged for backscratching for the first time.)
14	김학의 (Kim Hak Yi)[6] 구속영장 (Arrest warrant) 정준영 (Jeong Jun-young)	1. 김부겸·박상기, 문재인 대통령에 '장자연·김학의·버닝썬 사건' 보고 (Boo-kyum Kim and Sang-ki Park, reports about "Burning Sun Gate" to President Moon) 2. '버닝썬 유착' 의혹 윤 총경 출국 금지 ("Suspicion of collusion" on senior superintendent Yoon. Prohibition of departure on Yoon)
15	폭행 사건 (Assault case) 경리실장 (Chief accountant) 경리업무 (Accounting work)	1. '버닝썬' 경리담장자 돌연 출국 후 잠적("Burning Sun" accounting manager disappeared after departure) 2. 버닝썬 경리실장, 돌연 해외 출국 후 잠적…경찰 추적 중("Burningg Sun" accounting manager vanished after departure. Police force tracking him.)
16	린사모(Linsamo)[7] 김학의(Kim Hak Yi) 지창욱(Ji Chang-wook)[8]	1. "지창욱 버닝썬 게이트 관련 없어 린사모 사진은 참고용" '그것이 알고싶다' 측 해명 ("Chang-wook Ji is not related to Burning Sun gate" from TV program "그것이 알고싶다"[9]) 2. '그것이 알고싶다' 측 "지창욱, 버닝썬 린사모와 무관" (TV program "그것이

[6] He was vice minister of Ministry of Justice in South Korea. Accused for providing "Sexual favors".

[7] Investor of "Burning Sun" from Taiwan.

[8] Korean actor. As picture that was taken with Yi-Ju Lin was shown to public, there was suspicion that he was related to "Burning Sun".

[9] Korean TV program that probs about suspicious issues.

		알고싶다” says Chang-wook Ji is irrelevant with “Burning Sun gate”)
17	유착 의혹 (Suspicion of collusion) 로스쿨 교수들 (Professors at law school) 서강대 (Seo-gang university)	1. “버닝썬 동영상 잘리기 전에”… 서강대, ‘버닝썬 농담’ 로스쿨 교수들 조사 나서(Professor at law school jokes about “Burning Sun” sex tape) 2. 서강대 로스쿨 교수 “버닝썬 영상 짤리기 전에 봐라” 발언 논란(Seogang Universities law school professor, watch “Burning Sun” sex tape before it is gone)
18	김학의 (Hak-ui Kim) 유착 의혹 (Suspicion of collusion) 황하나 (Hana Hwang)[10]	1. 황하나 봐주기 수사 의혹, 경찰 이대로 괜찮나? ‘불신 쌓여’(Suspicion of collusion between police and Hana Hwang) 2. 오늘 국회 정보위 전체회의...김학의·버닝썬 질의 집중될 듯(May be meeting about “Burning Sun” and Hakui Kim at National Assembly Intelligence Committee)
19	멤버 승리 (Seung-Ri)[11] 정준영 (Jeong Jun-young) 삼합회 (Triad)	1. 로이킴도 단톡방에 음란물 유포… 참고인에서 피의자로(Roy Kim arrested for sending illegal obscene material. Witness to suspect) 2. 삼합회까지 연루? 버닝썬 수사 해외로 확장(Is Triad related? “Burnig Sun” investigation expands to overseas)
20	정준영(Jeong Jun-young) 연예인(Celebrity) 마약 투약 혐의 (Suspicion of drug administration)	1. ‘정준영 단톡방’→’황하나 마약’, 연예계만 떨면 그만?(Scandals about celebrities. Are celebrities sort of victims?) 2. ‘승리·정준영 게이트’ 이어 ‘황하나 게이트’ 열리나?(“Hana Hwang”gate right after “SeungRi, Jeongjun Young”gate?)
21	현직 경찰관 (Policeman in service) 중고차(Used car) 전직 경찰(ex-police)	1. ‘버닝썬 유착고리’ 전직 경찰, 현직 간부와 중고차 거래(Collusion ties between Burning Sun and police. (Through used car dealing)) 2. ‘버닝썬 유착’ 전직 경찰에게 중고차 산 경찰 입건(Police charged for collusion by used car.)

[10] Korean celebrity. Accused for drug dosage in club “Burning Sun”.

[11] Korean Celebrity. Establisher of “Burning Sun”.

22	정준영(Jeong Jun-young) 성관계 동영상(Sex tape) 김상교(Sangkyo Kim)[12]	1. 로이킴 10 일 경찰 출석, 음란물 카톡 유포 혐의 어떤 입장 내놓을까?(Roy Kim suspected for spreading illegal porn. What would he say?) 2. '정준영 단톡방'엔 무슨 일이 있었던 것일까(What has happened in their chat room?)
23	유리홀딩스 (Yuri Holdings)[13] 전원산업 (Jeonwon industry)[14] 가수 승리(SeungRi)	1. 버닝썬 횡령 의혹… 전원산업·유리홀딩스 압수수색(Suspicion of embezzlement of Burning Sun. Search and seizure on Jeonwon industry and Yuri holdings) 2. '승리 횡령 의혹' 전원산업 등 2 곳 압수수색…'린사모' 출석 요청(Suspicion of embezzlement on SeunRi. Search and seizure on Jeonwon industry and other places. Resquest of prosecution attendance on Linsamo)
24	이승현(SeungHyun Lee) 몽키 뮤지엄 (Monkey Museum)[15] 가수 승리(SeungRi)	1. 몽키뮤지엄 브랜드 사용료에 버닝썬 자금 수억원… 승리 추가 소환(Monkey Museum seems to have relationship with Burning Sun. SeungRi summoned for investigation) 2. 버닝썬 돈 2 억 몽키뮤지엄에 흘러가… 경찰, 횡령 여부 수사(Burning Sun's 200million won sent to Monkey Museum. Police investigation on embezzlement.)
25	마약 투약 혐의 (Suspicion of drug administration) 구속영장(Arrest warrant) 애나(Anna)	1. '버닝썬 마약' 이문호, 구속영장 발부...애나는 영장 기각 (Arrest warrant issued to CEO Lee for "Burning Sun drug case", rejection of arrest warrant on Anna) 2. 법원, '마약투' 버닝썬 MD 애나 구속영장 기각…"마약유통 혐의 소명 부족" (Court rejected arrest warrant on Anne for lack of suspicion)
26	JYP(JYP) 엔터테인먼트 업종 주가 (Stock price of Entertainment company)	1. 박진영, 이수만 제치고 연예인 주식부호 1 위에 올라(Park; CEO at JYP ent became stock rich among celebrities.)

[12] One who sued "Burning Sun" for assault. His accusation triggered "Burning Sun" gate.

[13] Company that invests on cosmetics, real estates or so. Seung-Ri was co-founder. There was suspicion that this company has relationship with "Burning Sun".

[14] Company that owns 42% shares of "Burning Sun".

[15] Club in Gangnam. This club was managed by Seung-Ri.

	가수 승리(SeungRi)	2. '버닝썬 여파' 이수만 488 억-양현석 322 억 주식재산 줄었다(CEO Lee and CEO Yang stock asset diminished because of "Burning Sun" gate's aftermath)
27	성폭행(Sexual assault) 정준영(Jeong Jun-young) 박한별(Han Byul Park)	1. 버닝썬 직원·정준영·최종훈 등 홍천 여행서 성폭행 정황…피해 여성 "남성들이 준 술 마시고 기억 끊겨"(Burning Sun's staff and other celebrities suspicion for sexual assault on their trip.) 2. 정준영·최종훈 홍천 여행서도 성폭행 의혹…女 "술 마신 뒤 기억 끊겨"(Junyoung Jeon and JongHoon Choi suspicion for sexual assault on their trip.)
28	마약 투약 혐의 (Suspicion of drug administration) 전원산업 (Jeonwon industry) 이문호(Mun Ho Lee)	1. '마약 투약' 버닝썬 이문호 대표 검찰 송치(CEO Lee summon to prosecution for Burning Sun drug administration) 2. '마약 혐의' 버닝썬 이문호·애나 검찰 송치(CEO Lee and Anna summon to prosecution for Burning Sun drug administration)
29	장자연 사건 (Jayeon Jang case)[16] VIP 등지 소각팀 (Incineration team)	1. "낮에 국가 기여했으니" VVIP 하룻밤 위해 투입된 가출 청소년들(Runway youth put into VVIP's sexual desire) 2. 버닝썬 핵심인물 '승리·윤총경·린사모' 수사 어떻게 돼가나 ('SeungRi, Senior superindtendent Yoon, Linsamo' are key members in Burning Sun gate.)
30	경찰 안팎 (Inside and outside the police) 경찰청(Police agency) 마약대응 조직 (Drug response organization)	1. 경찰청, 2020 년 마약범죄 전담기구 신설 추진(Police agency decided to establish new drug agency by 2020.) 2. 경찰청, 마약조직범죄과 신설 추진…수사인력 대폭 증원(Police agent decided to establish drug agent and reinforcement on investigating force.)
31	연예인들(Celebrities) 구속영장(Arrest warrant) 불법촬영(Illegal shooting)	1. 승리, 금주 내 구속영장 신청될까…경찰 "수사 막바지"(Would SeungRi's arrest warrant be registered by this week?)

[16] Suicide of Korean celebrity. There was suspicion that cause of suicide was forced "sexual favor" and sexual abuse.

		2. 경찰 "'성접대 의혹' 승리 이번 주 구속영장 신청 검토 중"(SurngRi's arrest warrant under review.)
32	전원산업 (Jeonwon industry) 정준영(Jeong Jun-young) 관련자(Related people)	1. '버닝썬 MD', 변호인 선임 지연 이유로 재판 또 연기(Trial suspended because "Burning Sun" MD could not hire lawyer.) 2. "승리, 이번 주 구속영장"…YG 수사 확대(SeungRi arrest warrant issued by this week. Also planning investigate YG; his entertainment company)
33	성매매(Prostitution) 외국인 마약사범 (Foreign drug criminal) 집중 단속 (Intensive crackdown)	1. 경찰 집중 마약단속으로 외국인 123 명 검거, 중국·동남아 출신 대부분(Police crackdown on drug. 123 people arrested. Most them are from east Asia.) 2. 커피에 과자에…경찰, 마약사범 집중단속 중 외국인 123 명 적발(123 foreigners caught for drug administration.)
34	유리 홀딩스 대표 (CEO at Yuri Holdings) 몽키뮤지엄 (Monkey Museum) 브랜드 음료 사용료 (Using brand beverage)	1. '버닝썬' 자금 횡령 혐의 승리 피의자 소환(SeungRi summoned as suspect for "Burning Sun" embezzlement) 2. 버닝썬 자금 횡령 혐의, 승리 피의자 신분으로 소환 조사(SeungRi Summoned as suspect for Burning Sun embezzlement.)
35	한효주(Han Hyo-joo)[17] JM솔루션(JM Solution)[18] 광고모델 (Advertising model)	1. 버닝썬 루머 논란에 한효주 측 "33명 네티즌 고소" (Hyo-joo Han charged 33 of Netizens for spreading rumors on "Burning Sun gate") 2. 한효주, 버닝썬 연관 루머 게시자 고소 (Hyo-joo Han accuses writers about rumors on "Burning Sun gate")

In the Table 10, each interval has at least one unique related keyword comparing with adjacent time intervals. From this, it can be inferred that unique keyword represents the new event of that interval. Also, first related keyword can be seen as mostly spotlighted keyword during that period.

However, there are some common keywords between intervals. For example, in time interval 8 and 9 in Table 10, both intervals contain "Suspicion of collusion" as related keyword. However, having common keyword is not problematic result since it indicates that

[17] Korean celebrity. She is model at JM Solution. There was suspicion about drug dosage at club "Burning Sun" party.

[18] Cosmetic company. JM Solution help party at "Burning Sun".

people are still interested in the past sub-issue along with new sub-issue. In this case both new keywords and old keywords can be on a list.

However, it is very clear that return result (including articles) is more helpful to understand whole event than scattered news articles. Algorithm at least removes unnecessary common news articles and organizes similar sub-issues which saves users' time and enhances users' understandability.

3.2. Result of Algorithm by Using Flow 2 at Figure 2 with Burning Sun Issue

When testing second example, starting and ending date of interval were included to enhance understanding. Also, length of interval was included for same purpose. Since second example uses flow 2 on Figure 2, it only returns 10 time intervals as return value. Table 11 is result of algorithm with given input mentioned below. In second example, exact time period of each interval was added in the return value. Time intervals with biggest eight weight sum were selected from 2nd to 34th time interval (In Table 10, there are 35 time intervals). First and last time intervals were included without ranking them to make full 10 time intervals of news "sub-issues". The reason why first and last time intervals were added is to show the starting point (or starting scandal) of issue and to keep the recent "sub-issue" updated.

Table 11. Result of timeline news algorithm with rephrasing option (Start date: 2019-01-01, keyword: "Burning Sun", Option:10).

Time Interval (Days)	Start End	Keywords	Title of article
1 (7)	01-29 02-04	폭행 사건 (Assault case) 성추행 (Sexual harassment) 성폭행 (Sexual assault)	1. 버닝썬 반박 "폭행 사건 인정..마약과 성추행 성폭행은 아냐" ("Burning Sun" reputes "Admits assault case but not sexual harassment and sexual assault") 2. '승리 클럽' 버닝썬, 디스패치 단톡방 보도에 "물뽕·성추행? 절대 동의 못해" ("Burning Sun"; belongs to Seung Ri, never agrees to usage of "GHB" and sexual harassment)
2 (7)	02-12	마약 투약 (Drug administration) 애나(Anna) 유착의혹(Suspicion	1. `버닝썬` 애나, 엑스터시 투약 적발..... 추방 결정 '불복' ("Burning Sun" Anna, injection of ecstasy…… objection to expulsion)

	02-18	of collusion)	2. 버닝썬 결국 계약 해지...클럽 MD '애나' 엑스터시 투약 의혹 (“Burning Sun” contract terminated…Anna(club MD) suspicion of ecstasy administration)
3 (7)	02-19 02-25	경찰관 (Police officer) 유착의혹 (Suspicion Of collusion) 압수수색 (Search and seizure)	1. 경찰 `버닝썬 유착 의혹` 은행계좌·통신 압수수색 (“Burning Sun” alleged collusion with police force. Search for seizure on bank accounts and communication record) 2. `버닝썬 유착` 계좌·통신 압수수색…… 현직 경찰관 여럿 포함돼 (“Burning Sun collusion” search for seizure on bank account and communication…… had several police officers in service)
4 (4)	02-26 03-29	성접대 (Sexual favor) 이문호 대표(CEO Lee) 사내이사(In-house director)	1. “버닝썬 이문호 대표 모발서 마약 양성반응 검출”..경찰, 승리 내사 착수 (Positive reaction on drug on CEO Lee’s hair. Police launch an investigation) 2. ‘승리 카톡’ 논란에 승리 “자진출두해 마약검사 받겠다”…경찰, 버닝썬 이문호 출국금지 (Seung Ri “hopes to voluntarily appear and get drug test”, CEO Lee forbidden to leave country)
5 (4)	03-02 03-05	이문호 대표 (CEO Lee) 유착의혹 (Suspicion of collusion) 마약 투약 (Drug administration)	1. '버닝썬' 이문호 대표, 10시간 경찰조사 뒤 새벽 귀가…"수사 협조했다" (“Burning Sun” CEO Lee got a police investigation for 10 hours and left at dawn) 2. 버닝썬 이문호 대표, 10시간 경찰 조사받고 귀가 (“Burning Sun” CEO Lee got 10 hours of police investigation and left)

6 (6)	03-10 03-15	정준영 (Jeong Jun-young) 유착의혹 (Suspicion of collusion) 경찰총장 (Police chief)	1. 민갑룡 경찰청장 “버닝썬 유착 의혹? 지위고하 막론 발본색원” (Kap-ryong Min a chief of police hopes eradicate all the suspicious police members related with “Burning Sun”) 2. 승리 카톡방 `경찰총장` 거론된 총경 "정준영 모른다" 참고인 조사 (Chief officer “A” mentioned on Seung Ri’s messenger. Investigation on potential witness)
7 (5)	03-17 03-21	김학의 (Kim Hak Yi) 구속영장 (Arrest warrant) 정준영 (Jeong Jun-young)	1. 김부겸·박상기, 문재인 대통령에 '장자연·김학의·버닝썬 사건' 보고 (Boo-kyum Kim and Sang-ki Park, reports about “Burning Sun Gate” to President Moon) 2. ‘버닝썬 유착’ 의혹 윤 총경 출국 금지 (“Suspicion of collusion” on senior superintendent Yoon. Prohibition of departure on Yoon)
8 (6)	03-23 03-28	린사모(Linsamo) 김학의 (Kim Hak Yi) 지창욱 (Ji Chang-wook)	1. "지창욱 버닝썬 게이트 관련 없어 린사모 사진은 참고용" '그것이 알고싶다'[19] 측 해명 (“Chang-wook Ji is not related to Burning Sun gate” from TV program “그것이 알고싶다”) 2. '그것이 알고싶다' 측 "지창욱, 버닝썬 린사모와 무관" (TV program “그것이 알고싶다” says Chang-wook Ji is irrelevant with “Burning Sun gate”)
9 (3)	04-17	마약 투약 혐의 (Suspicion of drug administration) 구속영장	1. '버닝썬 마약' 이문호, 구속영장 발부...애나는 영장 기각 (Arrest warrant issued to CEO Lee for “Burning Sun drug case”, rejection of arrest warrant on Anna)

[19] Korean journalism TV program that probs doubtful issues.

	04-19	(Arrest warrant) 애나(Anna)	2. 법원, '마약투약' 버닝썬 MD 애나 구속영장 기각…"마약유통 혐의 소명 부족" (Court rejected arrest warrant on Anne for lack of suspicion)
10 (1)	05-09 05-09	한효주 (Han Hyo-joo) JM솔루션 (JM Solution) 광고모델 (Advertising model)	1. 버닝썬 루머 논란에 한효주 측 "33명 네티즌 고소" (Hyo-joo Han charged 33 of Netizens for spreading rumors on “Burning Sun gate”) 2. 한효주, 버닝썬 연관 루머 게시자 고소 (Hyo-joo Han accuses writers about rumors on “Burning Sun gate”)

Table 11 is result of using issue “Burning Sun”. Since this algorithm with “option 10” only ranks the time intervals with same issue, relate keywords of time intervals can be also found in Table 10. Third time interval of Table 11 is same with fifth interval of Table 10. All other related keywords of each interval can be found in Table 10 since Table 11 is just rephrasing of Table 10. However, last time interval (most recent sub-issue) is different because test case2 was done on different date and recent news data has changed. Although this algorithm does not have visualization part, Burning Sun example would look like Figure 3 if hypothetical image of Figure 2 is applied.

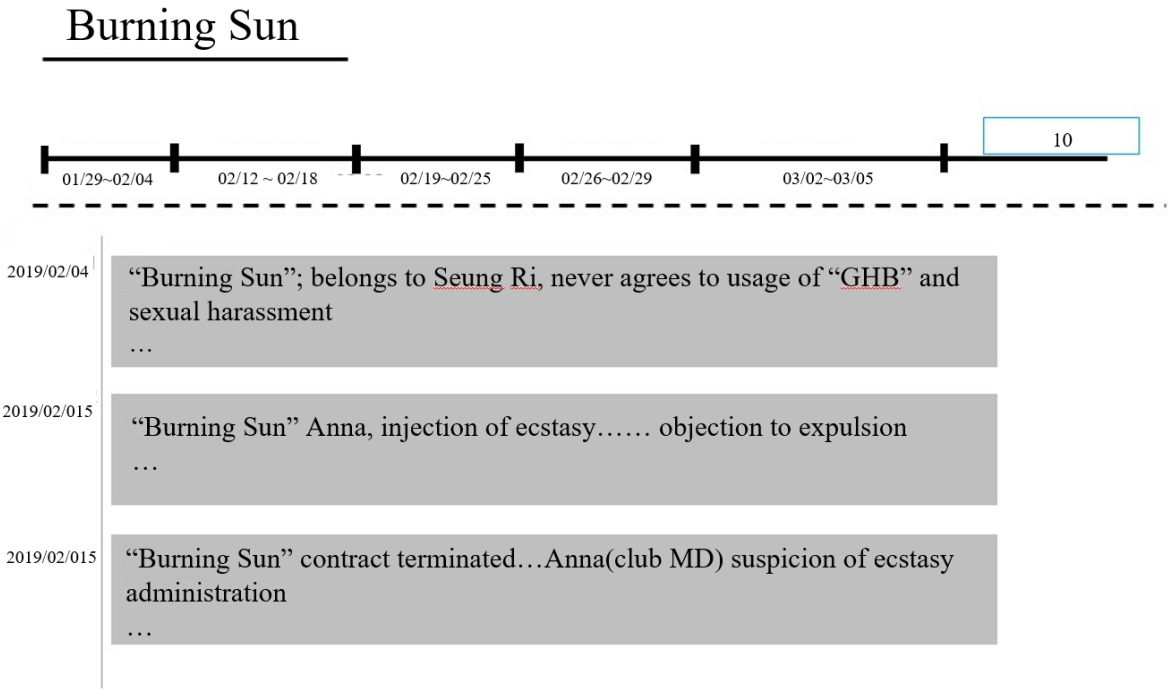

Figure 3. Hypothetical image of algorithm result using Burning Sun issue with rephrasing option of 10.

Other than that, effectiveness of result seemed pretty much similar with test case 1.

Despite effectiveness of both ways are similar, to compare characteristics between the two (One with option and one without option), One with option is much shorter than one without option so it would save time. However, some issues may be lost because algorithm tried to pick the most spotlighted sub-issues.

At this time, exact period was included in to output. It was obvious that showing exact period of each interval was much better since it shows the length of interval and show exact time point of a sub-issue.

3. Remark

Basic goal of this algorithm is to see the whole story line of given topic in a single webpage. To suit this need, algorithm should detect all the "sub-issues" within a given timeline. There was no time interval that all three related keywords were same with other time interval, which indicates successful result of proposed algorithm. In addition to that, we are successful to separate sub-issues even with the exact same top keyword. For example, 20th and 22nd time intervals shown in Table 10 shares the exact same top keyword but successfully identified different sub-issues. The strength of this algorithm is that it can sensitively reflects what people were interested at certain time period. Also, it helps to understand the whole flow of an event without searching or looking at all the news all along.

There are some key differences in two case studies. The former generally takes more time to read the produced result because the result contains more time intervals so in turn more details. The later, more compact form of the former case is more helpful to understand the given issue because it shows the exact dates of each time intervals along with keywords. However, we could choose to include the dates in the former case as well. The main difference is the number of subintervals, which can be controlled by the option in the proposed algorithm.

As time progresses, it is possible that neatly summarized news written by someone may show up. However, the algorithm cannot catch such summarizing news because summary news contains past keywords only. Therefore, it is highly likely that summary news and timeline does not have related keywords in common. Hence, this algorithm has strength in originality since it detects news articles that are posted for the first time.

Reference

Eunyoung, Kang 2010, *A Study on the Cases of Applying Interactive Storytelling on the Web Site*, Department of Media Broadcasting Graduate School of Information Sciences 2010.2

Haeyeop, Song and Jay, Yang 2017, *Online News Portal Service and changes in News Distribution: Big data Analysis of Naver News in 2010-2017*, Korean Journal of journalism &Communication Studies 61(4)

Joan, MacDonald 2019, *How The Burning Sun Scandal Affected Korea's Drama World*, *Forbes*

Ji-Yeon, Kim and Jae Young, Yun 2015, *A Study on Interactive Stroytelling Methods for Online News*, International Design Conference of KSDS and ADADA with Cumulus

Kyungmo, Kim 2012, *Online News Production in the New Journalism Environment Between Tradition and Change*, Journal of communication research 49(1)

Wi-Geun, Kim 2014, *The Influence of Portal Site News Services on Online Journalism in Korea: The Structural Transformation or the Power Change in the News Distribution*, Korean Journal of Communication & Information (2014.5)

Yeo-Kwang, Yoon 2014, *A Study on Contents Curation of Portal Sites*, Journal of the Korea Entertainment Industry Association 8(4)