Journal of
Information Technology & Software Engineering

Alsuhaymi, J Inform Tech Softw Eng 2017, 7:3
DOI: 10.4172/2165-7866.1000204

Research Article Open Access

Provenance Detection of Online News Article

Ruba Ali Alsuhaymi*

Department of Software Engineering, Prince Sultan University, Riyadh, Saudi Arabia

Abstract

At present, with the current wide spread of information on the social media, the recipient or the researcher needs more details about the received information or spread, including the provenance. With the current explosion of the news websites, there is a question of credibility of news articles on the internet. It is important to know whether the news is correct or not. This paper focuses on identifying the provenance of news articles. Also, trace the provenance of news articles often to see where did the first publication of such news appear. Is the news publication true (the credibility of the news), or is the news quoting from the provenance of the news on the news website or is plagiarism and redistributed on news websites on the Internet? In this paper, we will answer these questions through the design and implementation of two techniques Google Search API and Google Custom Search that will define the provenance of news articles through the technique Topic Detection and Tracking (TDT). Therefore, verifies the proposed technical quality in terms of performance metrics through several different experiments. Based on these experiments and tests it were discovered that the technique Google Search API is better performance than Google Custom Search in detecting the provenance of news articles. The Google Search API is the best technique, depending on the user satisfaction, the time it takes to view the results and the accuracy and validity. So, the result of the Google Search API is 90% while Google Custom Search 70%.

Keywords: Detection; News articles; Provenance; Plagiarism

Introduction

Nowadays, with the rapid growth of the Internet and the increasing amount of information on the Internet, and increase the number of news articles and news websites, huge amount of news articles that are published every day. So, the users need to know more details of information about the news articles published, including the provenance and the personal attributes of the user, like name, sex, education, location and race [1]. Thus, the challenges and the main purpose for the users is the capability for tracing and detecting the reliability and truth of information from among thousands of results. Through helping the readers to ensure the truthfulness of the news (the credibility of the news) through reading the news from news websites reliable. Also, check if the news website is plagiarized the news article and redistributed it on the news websites on the Internet or quoted the news article from the news website provenance. Finally, identify whether a news website is the first publication of such news. While the provenance is not limited to determining the news articles, the heritage of the artwork, archeology, paleontology, archives, manuscripts, printed books and computing science, is valid too many areas to find out their provenance facts.

Provenance is the history or chronology of the ownership of a valued object or work of arts or literature or location of historical objects. The provenance of an article can be defined as the information about the entities, activities and people involved in the production of a piece of data, and such can be used as information to evaluate the quality, reliability and confidence of the data [2]. Provenance (also referred to as lineage, pedigree, parentage, genealogy, and filiation) [3]. It can be described the provenance at different intervals depended on where it was used. Buneman et al. [4] definition the provenance of the data through the database systems, as well as describing the provenance of data and processes that reach into the database. Greenwood et al. [5] widen the definition of the provenance by recording workflow process through experience.

Plagiarism is theft of intellectual property, whether research or art or invention, etc. The reason of plagiarism is the easy access to web pages and databases, so the plagiarism is a big problem for publishers, researchers [6].

The benefit of this study is tracing and detecting the process of monitoring a stream of news articles in order to find those news articles that track (or discuss) the same event. The objective is to focus on detecting and organization the provenance of the news articles. As a result, the model will be developed to help individuals and organizing to define the provenance of news articles on any news websites also, support all languages such as Arabic and English etc. On the other hand, the research contributes to society by achieving the following objectives, help the reader to identify whether a news website is the first publication of such news, helping the reader to ensure the truthfulness of the news (the credibility of the news) through reading the news from news websites is reliable, or detect and determine whether the news article was quoting from the provenance or was plagiarism and redistributed on news websites on the Internet.

This paper describes the techniques that detect the provenance of news articles on the internet. The rest of this paper is organized as follows: Section 2 describes the related work. Section 3 describes the methods. Section 4 presents the experiments of the system and the results. Section 5 discusses the results of the system. Finally, give the conclusions and discussions of future work in section 6.

Related Work

Wylot et al. [7] the TripleProv is an approach used to collect data, storing, organizing, and tracking provenance information and then displaying the provenance. It can enter the provenance specification of the data one wants to use in order to reach the answer. Such as, if interest in the articles is about "Corona Virus" but you want the answer to come from provenance attributed to the "CNN News ".

Nallapati [8] presents a new approach to relate detection for Topic

***Corresponding author:** Ruba Ali Alsuhaymi, Department of Software Engineering, Prince Sultan University, Rafha St, Salah Ad Din, Riyadh 12435 11586, Saudi Arabia, Tel: + 0555463405; E-mail: ruba.alsuhaymi@gmail.com

Received May 04, 2017; **Accepted** July 19, 2017; **Published** August 04, 2017

Citation: Alsuhaymi RA (2017) Provenance Detection of Online News Article. J Inform Tech Softw Eng 7: 204. doi: 10.4172/2165-7866.1000204

Detection and Tracking (TDT) that will define the provenance of news articles. Topic Detection and Tracking (TDT) is a form of event-based information organization [9]. The goal of Topic Detection and Tracking (TDT) is gathering the news to groups to discuss a single topic.

Nies et al. [10] the resulting derivations are structured in the Provenance Data Model (PROV-DM), which is the discovery of provenance of news articles. With this approach, it was discovered that the provenance of news articles by 73% out of 410 news articles, and with 68% accuracy. The Provenance Data Model (PROV-DM), is being developed by the World Wide Web Consortium (W3C1) Provenance Working Group.

In Chang et al. [11] the focus is to design and implementation of an application that summarizes and tracks algorithms for Chinese news, and give lists of news that depends on the timestamp in orders to guarantee ease of understanding. Also, it has implemented Term Frequency- density (TF–density) algorithm work to weight the term list for each event. Also, it makes a word vector for each event by utilizing the term information table. Term Frequency- density (TF–density) algorithm that is proposed and compared with the algorithms Term Frequency– Inverse Document Frequency (TF-IDF) and Term Frequency– Inverse Word Frequency (TF-IWF).

The research provides an experimental analysis and focuses on improving means of summarizing a group of news articles to help readers to understand topics quicker. It is limited to tracking Chinese language news articles.

Bea and Claire [12] describes the Synergetic Content Creation and Communication (SYNC3) project for development of a system for tracking news articles. An event is defined in the news by using TDT techniques to know a particular time and place of event. The system tracks the provenance of news articles. Then these groups are handled by labeling and extracting the temporal and geographical relations between the events.

It is used to develop 12,547 documents from nine different news provenance "(AP, BBC, CNN, NYT, Reuters, Ria Novosti, USA TODAY, WP and Xinhua)" from the date 20 May to June 3, 2009 [12].

This study focused to labeling of the news article, with labeling the topic and the foundations of the temporal, spatial news articles, and provide some preliminary statistics on the data. This allows determining the suitability of the labels news articles as well as dates, locations, and adjust the accuracy of the labeling process [12].

Kamalpreet, Balkrishan [13] discussed a way to detect plagiarism through the crawl service provided by the custom search engine API using semantic technology. This approach is searching through a focus on keywords.

The verification of plagiarism in documents, through measuring the percentage of similarity and matching of a string of the document on number of n-grams participate between different documents. The results of the proposed method are a quick and efficient and accurate.

Methods

To detect the provenance of the news articles on the news websites on the Internet. The method proposed for detecting the provenance of news articles is Topic Detection and Tracking technique, through using Google Custom Search or Google Search API. Google Custom Search is a platform offered by Google that allows web developers to offer customized information in web search results. Also, classified and organize the queries and create customized search engines based on Google search [14,15]. Google Search API is used to provide keywords for search engine Google and get the retrieval results from Google, and the combination of these two things can help users to find the information needed to better search [16].

The method consists of four interrelated phases. The first phase is search for the news articles on all the news websites, the second is printing the first published news article, the third is determined the type of news articles published by relying on the first news article published, and the final is displaying the results. These phases depend on each other.

In the first phase, enables the user to enter any news article, whether a word or several words or a phrase, or title of a news article, or even the full text of the news article in the text box; to search for all the similar news articles in the news websites. The second phase, displays the title of the news article and news websites link, date and time of the publication of the news article and the type of news article. Also, know the first publisher of the news, by searching for the oldest date among all publications of the same news article, considering the oldest date for each news article in the news website that is a first publisher the news article. The third phase, determined the type of news articles published, through compare the first news publication with other news articles, to detect if the news was quoted or plagiarized the news article, and it calculates the percentage of similarity and thus, printing "Plagiarism" if the percentage of copies and quote is great and not refer to the news provenance, but print "Quote this news article from the provenance of news website " when refer to the news provenance, or "Not related to this title" if this title not related to the title search, or print "This title written by this website" to tell the user this title copied from this website. Also, the search results are organized by the date and time.

Final phase, display the results and print the website name (link) that first news website published the news article, and the title of the news; to make it easier for the reader or researcher to find the provenance of news. Also, the user can compare the first news article enter with other news published on news websites, or compare the first news article was published with news articles published on other news websites, as shown in Figure 1.

On the other hand, are validated model proposed by experiment of two different techniques are Google Custom Search and Google Search API, in next section.

Experimental Results

In this section, will display the experiments and their results, by dividing the experience into two parts: the first experiment was applied Google Search API to detect the provenance of news articles in all the news websites in English and Arabic language. The second experiment was applied Google Custom Search service to detect the provenance of news articles to certain news websites in English and Arabic language.

In each experiment, will be the experiment the program by the 10 participants. To confirm and validate these experiments, it should be measured through using software metric factors, particularly factors that mean "how well does the tool run", factors of the product are Effectiveness, Efficiency and Satisfaction users [17,18]. So, after each experiment of the program, the participant will respond to the questions that display in Table 1 for first experiment and Table 2 for the second experiment. To determine the outcome or feedback of the experiment the program and how effective the program. In Tables 1 and 2 can show the usability standards during the user experiment this task. These standards are task, completion of the involved tasks, and the time to complete tasks, and satisfaction from the experiment. All these standards to measure based on user feedback. Thus, give the results of the completion of tasks or the time consumed to complete tasks from the user view which evaluates the experiment.

First experiment

In this experiment, we used the Google Search API to search in any news websites. The results of the questionnaire of the experiment are shown in Table 1. Also, the results of the experiment will show in Figures 1 and 2.

Advantages:

1. It allows to search in all the news websites only and support all languages.
2. The program arranged news articles by the date and time from the oldest to the newer and vice versa.
3. The software will also not allow the ads to appear in search results.
4. Allows viewing 100 results.
5. Can click on the link of news to read the full details.

Limitation:

1. Search in any news websites that may be no reliability for news articles.

Second experiment

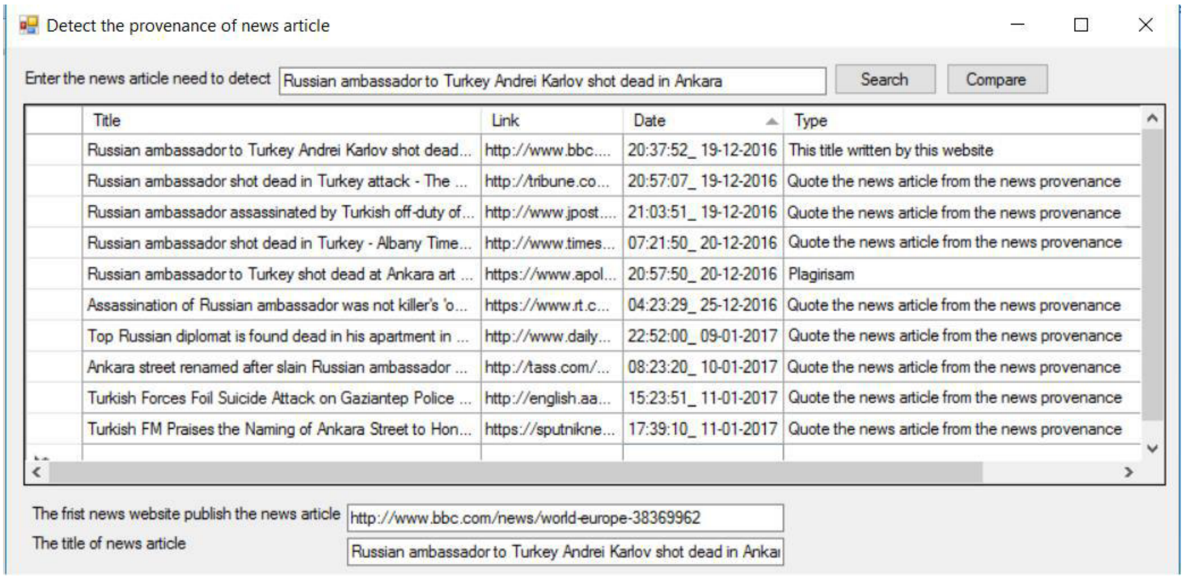
In this experiment, we used Custom Search Google to search for news websites selected such as: BBC, CNN, CNBC, SPY, Aljazeera, Reuters, New York Times, Huffington Post, Al-Arabiya, Daily Mail. These websites were chosen based on a survey was published through social media. It received 73 responses, 43 responses by were female and 30 were male. Also, it ensures 3 people of PhDs and 13 Masters and 37 Bachelor's and 20 High school graduate. The results of the questionnaire of the experiment is shown in Table 2. Also, the results of the experiment will show in Figures 3 and 4.

Advantages:

1. It allows to search in any news websites selected before and support all languages.
2. The program arranged news articles by the date and time from the oldest to the newer and vice versa.
3. The software will also not allow the ads to appear in search results.
4. The websites that are being searched for news websites and has a reliable, professional standard for the news. Also, does not search in blog or any web pages on the Internet are not interested for the news articles and unreliable.

Usability Standards	Description				
Effectiveness	Task completion:				
	How describe the difficulty or easy to achieve these tasks?				
	Task	Completed successfully	Completed with difficulties	Failed	Not used used
	Enter the news article to be searched	100%			
	Click the search button	100%			
	Shows the results	90%	10%		
	Click compares button if the user wishes to compare the news article published with the first news article published	70%	10%		20%

Table 1: The usability results for the first experiment.



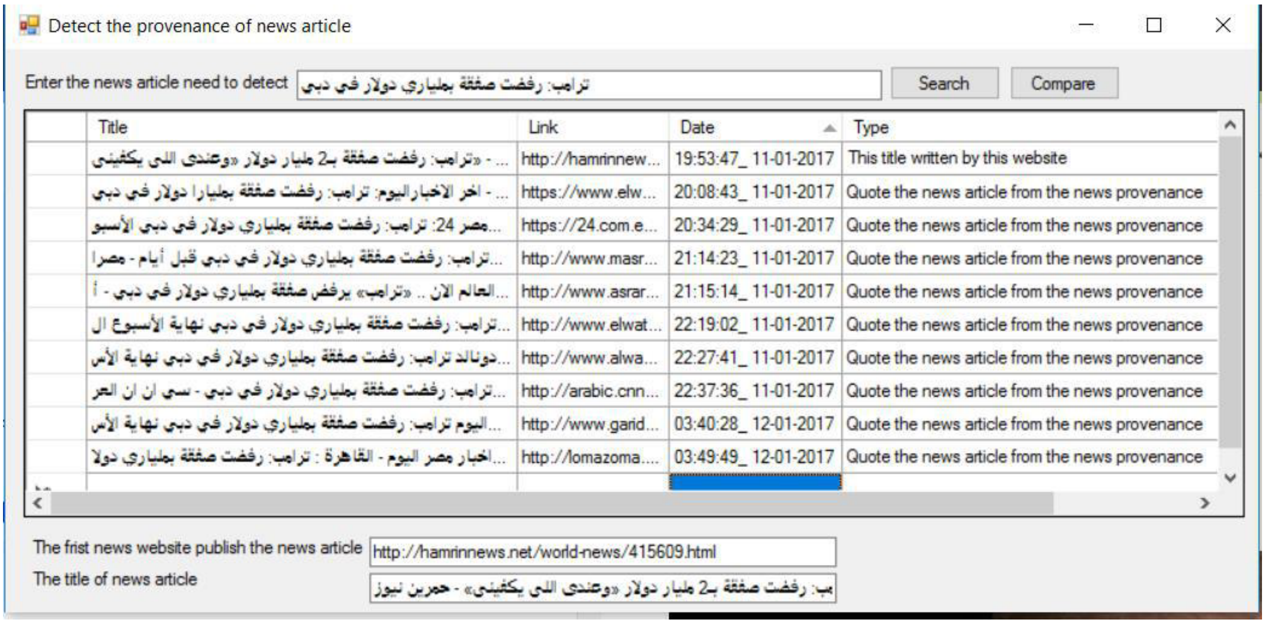
Enter the news article need to detect: Russian ambassador to Turkey Andrei Karlov shot dead in Ankara

Title	Link	Date	Type
Russian ambassador to Turkey Andrei Karlov shot dead...	http://www.bbc....	20:37:52_ 19-12-2016	This title written by this website
Russian ambassador shot dead in Turkey attack - The ...	http://tribune.co...	20:57:07_ 19-12-2016	Quote the news article from the news provenance
Russian ambassador assassinated by Turkish off-duty of...	http://www.jpost....	21:03:51_ 19-12-2016	Quote the news article from the news provenance
Russian ambassador shot dead in Turkey - Albany Time...	http://www.times...	07:21:50_ 20-12-2016	Quote the news article from the news provenance
Russian ambassador to Turkey shot dead at Ankara art ...	https://www.apol...	20:57:50_ 20-12-2016	Plagirisam
Assassination of Russian ambassador was not killer's 'o...	https://www.rt.c...	04:23:29_ 25-12-2016	Quote the news article from the news provenance
Top Russian diplomat is found dead in his apartment in ...	http://www.daily...	22:52:00_ 09-01-2017	Quote the news article from the news provenance
Ankara street renamed after slain Russian ambassador ...	http://tass.com/...	08:23:20_ 10-01-2017	Quote the news article from the news provenance
Turkish Forces Foil Suicide Attack on Gaziantep Police ...	http://english.aa...	15:23:51_ 11-01-2017	Quote the news article from the news provenance
Turkish FM Praises the Naming of Ankara Street to Hon...	https://sputnikne...	17:39:10_ 11-01-2017	Quote the news article from the news provenance

The frist news website publish the news article: http://www.bbc.com/news/world-europe-38369962

The title of news article: Russian ambassador to Turkey Andrei Karlov shot dead in Ankai

Figure 1: Comparison of the news article that entered with other news articles in English.



Detect the provenance of news article

Enter the news article need to detect: ترامب: رفضت صفقة بملياري دولار في دبي

Title	Link	Date	Type
... - «ترامب: رفضت صفقة بـ2 مليار دولار «وعندى اللى يكفينى	http://hamrinnew...	19:53:47_ 11-01-2017	This title written by this website
... - اخر الاخباراليوم: ترامب: رفضت صفقة بمليارا دولار في دبي	https://www.elw...	20:08:43_ 11-01-2017	Quote the news article from the news provenance
...مصر 24: ترامب: رفضت صفقة بملياري دولار في دبي الأسبو	https://24.com.e...	20:34:29_ 11-01-2017	Quote the news article from the news provenance
...ترامب: رفضت صفقة بملياري دولار في دبي قبل أيام - مصرا	http://www.masr...	21:14:23_ 11-01-2017	Quote the news article from the news provenance
...العالم الان .. «ترامب» يرفض صفقة بملياري دولار في دبي - أ	http://www.asrar...	21:15:14_ 11-01-2017	Quote the news article from the news provenance
...ترامب: رفضت صفقة بملياري دولار في دبي نهاية الأسبوع ال	http://www.elwat...	22:19:02_ 11-01-2017	Quote the news article from the news provenance
...دونالد ترامب: رفضت صفقة بملياري دولار في دبي نهاية الأس	http://www.alwa...	22:27:41_ 11-01-2017	Quote the news article from the news provenance
...ترامب: رفضت صفقة بملياري دولار في دبي - سي ان ان العر	http://arabic.cnn...	22:37:36_ 11-01-2017	Quote the news article from the news provenance
...اليوم ترامب: رفضت صفقة بملياري دولار في دبي نهاية الأس	http://www.garid...	03:40:28_ 12-01-2017	Quote the news article from the news provenance
...اخبار مصر اليوم - القاهرة : ترامب: رفضت صفقة بملياري دولا	http://lomazoma....	03:49:49_ 12-01-2017	Quote the news article from the news provenance

The frist news website publish the news article: http://hamrinnews.net/world-news/415609.html

The title of news article: ب: رفضت صفقة بـ2 مليار دولار «وعندى اللى يكفينى» - حمرين نيوز

Figure 2: Comparison of the provenance of news article that published with other news articles in Arabic.

Efficiency	Consumed time:				
	What amount of time it took to achieve these tasks?				
	Task	**Very little time between (5-15 seconds)**	**Medium time between (16 seconds -1 minutes)**	**Too much time (more than one minutes)**	**Not used**
	Enter the news article to be searched	100%			
	Click the search button	100%			
	Shows the results	90%	10%		
	Click compares button if the user wishes to compare the news article published with the first news article published	70%	10%		20%
Satisfaction	**Questions:**				
	Questions	**Very satisfied**	**Normal satisfied**	**Unsatisfied**	
	How satisfied while using this application for the detection the provenance of news articles?	90%	10%		

Table 2: The usability results for the second experiment.

5. Allows viewing 100 results.
6. Can click on the link of news to read the full details.

Limitations:

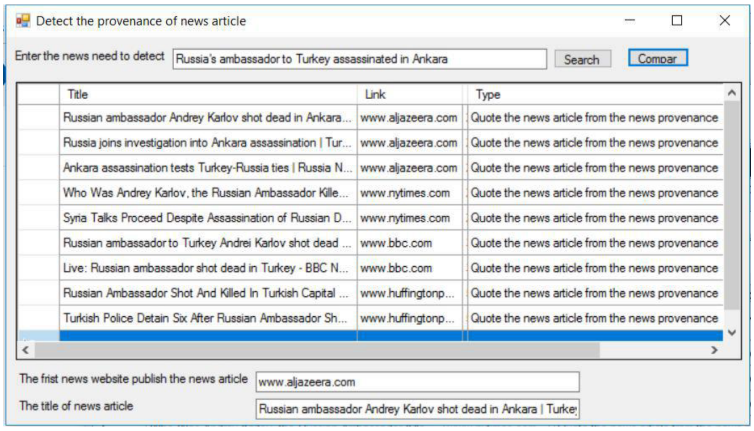
1. Sometimes it does not display the time and date.
2. Sometimes it does not display all the results accurately.

Finally, through two experiments using Google Search API and Google Custom Search. The results of the Google Search API are better than Google Custom Search to detect the provenance of news articles. The Google Search API best technique, depending on the user satisfaction and the time it takes to view the results and the accuracy and validity. Show in the figure each experiment is recorded and discussed. Therefore, the use of the program reduces the time and effort in the search for news articles and classified according to the older for the date and time and with accuracy and fast. Also, display the time and date of the news article and determine the type of news article is a plagiarism, or was excerpted the news article from the provenance of the news on news website, or is it modified or not has any related to the title of the news article.

Discussion

There are two experiments involving 10 participants, and measured the usability to use and the time it takes for each task and finally the satisfaction the participant for the experiments. When looking at the results generally of experiments that presented in Tables 1 and 2, therefore, increased rate of satisfaction to achieve the desired usability of the terms: the ability to complete all tasks successfully, take a little time to finish the tasks, and satisfaction to use the program. In Table 3 show the usability results for different experiments.

Finally, through two experiments using the Google Search API and Google Custom Search. The results of Google Search API are better than Google Custom Search to detect the provenance of news articles. The Google Search API best technique, depending on the user satisfaction and the time it takes to view the results and the accuracy and validity.



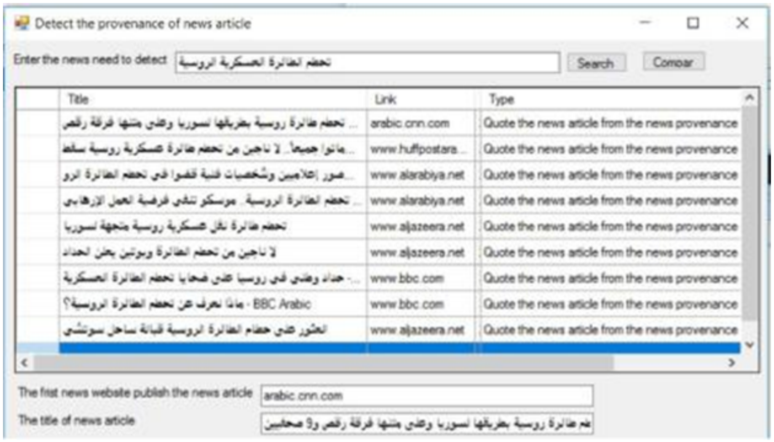
Enter the news need to detect: Russia's ambassador to Turkey assassinated in Ankara

Title	Link	Type
Russian ambassador Andrey Karlov shot dead in Ankara...	www.aljazeera.com	Quote the news article from the news provenance
Russia joins investigation into Ankara assassination | Tur...	www.aljazeera.com	Quote the news article from the news provenance
Ankara assassination tests Turkey-Russia ties | Russia N...	www.aljazeera.com	Quote the news article from the news provenance
Who Was Andrey Karlov, the Russian Ambassador Kille...	www.nytimes.com	Quote the news article from the news provenance
Syria Talks Proceed Despite Assassination of Russian D...	www.nytimes.com	Quote the news article from the news provenance
Russian ambassador to Turkey Andrei Karlov shot dead ...	www.bbc.com	Quote the news article from the news provenance
Live: Russian ambassador shot dead in Turkey - BBC N...	www.bbc.com	Quote the news article from the news provenance
Russian Ambassador Shot And Killed In Turkish Capital ...	www.huffingtonp...	Quote the news article from the news provenance
Turkish Police Detain Six After Russian Ambassador Sh...	www.huffingtonp...	Quote the news article from the news provenance

The frist news website publish the news article: www.aljazeera.com

The title of news article: Russian ambassador Andrey Karlov shot dead in Ankara | Turke

Figure 3: Comparison of the news article that entered with other news articles in English.



Enter the news need to detect: تحطم الطائرة العسكرية الروسية

Title	Link	Type
تحطم طائرة روسية بطريقها لسوريا وعلى متنها فرقة رقص ...	arabic.cnn.com	Quote the news article from the news provenance
ماتوا جميعاً.. لا ناجين من تحطم طائرة عسكرية روسية سقط...	www.huffpostara...	Quote the news article from the news provenance
صور إعلاميين وشخصيات فنية قضوا في تحطم الطائرة الرو...	www.alarabiya.net	Quote the news article from the news provenance
تحطم الطائرة الروسية.. موسكو تنفي فرضية العمل الإرهابي ...	www.alarabiya.net	Quote the news article from the news provenance
تحطم طائرة نقل عسكرية روسية متجهة لسوريا	www.aljazeera.net	Quote the news article from the news provenance
لا ناجين من تحطم الطائرة وبوتين يعلن الحداد	www.aljazeera.net	Quote the news article from the news provenance
حداد وطني في روسيا على ضحايا تحطم الطائرة العسكرية - ...	www.bbc.com	Quote the news article from the news provenance
ماذا نعرف عن تحطم الطائرة الروسية؟ - BBC Arabic	www.bbc.com	Quote the news article from the news provenance
العثور على حطام الطائرة الروسية قبالة ساحل سوتشي	www.aljazeera.net	Quote the news article from the news provenance

The frist news website publish the news article: arabic.cnn.com

The title of news article: طم طائرة روسية بطريقها لسوريا وعلى متنها فرقة رقص و9 صحفيين

Figure 4: Comparison of the provenance of news article that published with other news articles in Arabic.

Task	**Google Search API**		**Google Custom Search**	
	Completed successfully	Completed in Very little time between (5-15 seconds)	Completed successfully	Completed in Very little time between (5-15 seconds)
Enter the news article to be searched	100%	100%	100%	100%
Click the search button	100%	100%	100%	100%
Shows the results	90%	90%	80%	80%
Click compares button if the user wishes to compare the news article published with the first news article	70%	70%	70%	70%

published				
How satisfied while using this application for the detection the provenance of news articles	90%		70%	

Table 3: The usability results for different experiments.

As shown in Table 3.

Conclusion and Future Work

In this paper, we have investigated the detection of the provenance of news articles in various news websites on the Internet through developing a program for detection the first spread of the news article and the reliability of the news article. Also, determined if quotes the news article from the provenance of the news on the news website, or the news article has been plagiarized and redistributed on news website on the Internet, or not has any related to the news article. Finally, if the news websites where publication credible and correct a news article or not. They also discussed the relevant offering approaches, models and techniques related to the provenance in the literature review. As well as, display the projects, tools and theories about the provenance and that are related to the topic of the paper. It has been used Google Search API due to its efficacy in detecting the provenance of news articles. It involves detecting the provenance of news articles into four interrelated phases of the process. The first phase is search for the news articles on all the news websites, the second is printing the first published news article, the third is determined the type of news articles published by relying on the first news article published, and the final is displaying the results. These phases depend on each other. On the other hand, are validated model proposed by experiment of two different techniques are Google Custom Search and Google Search API. It has been the conclusion that the technique Google Search API is best through factor ease of use and user satisfaction with the results.

For the future works, will focus to enhance the results through increase the accuracy of the results. Furthermore, it can detect the provenance of the news articles through the inclusion of the photo or video or voice for news article.

References

1. Gundecha P, Ranganath S, Feng Z, Liu H (2013) A tool for collecting provenance data in social media. 19th ACM Special Interest Group on Knowledge Discovery in Data (SIGKDD), International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, Illinois, USA. pp: 1462-1465.
2. Groth P, Moreau L (2013) PROV-Overview An Overview of the PROV Family of Documents. World Wide Web Consortium.
3. Simmhan YL, Plale B, Gannon D (2005) A Survey of Data Provenance Techniques. Computer Science Department, Indiana University.
4. Buneman P, Khanna S, Tan WC (2001) Why and Where: A Characterization of Data Provenance. Lecture Notes in Computer Science, pp: 316-330.
5. Greenwood M, Goble CA, Stevens RD, Zhao J, Addis M, et al (2003). Provenance of e-Science Experiments-experience from Bioinformatics. UK e-Science All Hands Meeting 2003, East Midlands Conference Centre, Nottingham, pp: 223-226.
6. Maurer HA, Kappe F, Zaka B (2006) Plagiarism - A Survey. Journal of Universal Computer Science 12: 1050-1084.
7. Wylot M, Cudré-Mauroux P, Groth PT (2015) A Demonstration of TripleProv: Tracking and Querying Provenance over Web Data. International Conference on Very Large Data Bases (PVLDB) 8: 1992-2003.
8. Nallapati R (2003) Semantic Language Models for Topic Detection and Tracking. Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology. Proceedings of the HLT-NAACL 2003 Student Research Workshop 3: 1-6.
9. Makkonen J, Ahonen-Myka H, Salmenkivi M (2003) Topic Detection and Tracking with Spatio-Temporal Evidence. ECIR'03 Proceedings of the 25th European conference on IR research ECIR, Pisa, Italy. pp: 251-265.
10. Nies TD, Coppens S, Deursen DV, Mannens E, De Walle RV (2012) Automatic Discovery of High-Level Provenance Using Semantic Similarity. Provenance and Annotation of Data and Processes - 4th International Provenance and Annotation Workshop. pp: 97-110.
11. Chang HT, Liu SW, Mishra N (2015) A Tracking and Summarization System for Online Chinese News Topics. Aslib Journal of Information Management. Emerald Group Publishing Limited 67: 687-699.
12. Alex B, Glaire C (2010) Labelling and Spatio - Temporal Grounding of News Events. Proceedings of the Workshop on Computational Linguistics in a World of Social Media at NAACL 2010, Los Angeles, USA. pp: 27-28.
13. Sharma K, Jindal B (2016) An improved Online Plagiarism Detection Approach for Semantic Analysis using Custom Search Engine. 3rd International Conference on Computing for Sustainable Global Development (INDIACom).
14. Gulli A, Signorini A (2005) The Indexable Web is More than 11.5 Billion Pages. WWW'05: Special Interest Tracks and Posters of the 14th International Conference on World Wide Web. pp: 902-903.
15. Allauddin M, Azam F (2011) Service Crawling using Google Custom Search API. International Journal of Computer Applications 34: 10-15.
16. Yunpeng C, Peng T, Shihong L, Sufen S (2010) Study of Agricultural Search Engine Based on FAO Agrovoc Ontology and Google API. Proceedings 2010 World Automation Congress (WAC), Japan. pp: 439-444.
17. Galin D (2004) Software Quality Assurance: From Theory to Practice. Pearson, England.
18. Cavano JP, McCall JA (1978) A Framework for the Measurement of Software Quality. ACM SIGMETRICS Special Interest Group on Performance Evaluation Review 7: 133-139.