

# Term-Based Approach for Linking Digital News Stories

Muzammil Khan<sup>1</sup>(✉), Arif Ur Rahman<sup>2</sup>, and Muhammad Daud Awan<sup>1</sup>

<sup>1</sup> Department of Computer Science, Preston University, Islamabad, Pakistan  
muzammilkhan86@gmail.com, drdaudawan@preston.edu.pk

<sup>2</sup> Department of Computer Science, Bahria University, Islamabad, Pakistan  
badwanpk@gmail.com

**Abstract.** The World Wide Web has become a platform for news publication in the past few years. Many television channels, magazines and newspapers have started publishing digital versions of the news stories online. It is observed that recommendation systems can automatically process lengthy articles and identify similar articles to readers based on a predefined criteria i.e. collaborative filtering, content-based filtering approach. The paper presents a content-based similarity measure for linking digital news stories published in various newspapers during the preservation process. The study compares similarity of news articles based on human judgment with a similarity value computed automatically using common ratio measure for stories. The results are generalized by defining a threshold value based on multiple experimental results using the proposed approach.

**Keywords:** Linking news stories · Similarity measures  
Text processing

## 1 Introduction

The advanced technologies and proliferation of the Internet attract news readers to read online news from multiple sources and get the desired information. It is not humanly possible to browse through a huge information space for related information items. The amount of news article releases has grown rapidly and for an individual it is cumbersome to browse through all online sources for relevant news articles. Information retrieval techniques help in searching through the vast information spaces up to some extent and recommendation systems have emerged to respond the challenge by providing users the information which matches their needs either by their preferences or by content similarity among the news items. Each online news provider tries to handle their news articles and use some mechanisms to recommend similar news to the news readers.

The news generation in the digital environment is no longer a periodic process with a fixed single output like printed newspaper. The news are instantly generated and updated online in a continuous fashion. However, because of different

reasons like the short lifespan of digital information and speed of generation of information, it has become vital to preserve digital news for the long-term.

Digital preservation includes various actions to ensure that digital information remains accessible and usable, as long as they are considered important [3]. Libraries and archives preserve newspapers by carefully digitizing collections as newspapers are a good source of knowing history. Many approaches have been developed to preserve digital information for the long term [5, 15]. The lifespan of news stories published online vary from one newspaper to another, i.e. from one day to a month. Though, a newspaper may be backed-up and archived by the news publisher or national archives, in the future it will be difficult to access particular information published in various newspapers about the same news story. The issues become even more complicated if a story is to be tracked through an archive of many newspapers, which require different access technologies.

The Digital News Story Extractor (DNSE) is a tool developed to facilitate the extraction of news stories from the online newspapers and migrate to a normalized format for preservation using Digital News Stories Preservation (DNSP) framework [8]. The normalized format also includes a step to add metadata in the Digital News Stories Archive (DNSA) for future use [9]. To facilitate the accessibility of news articles preserved from multiple sources, some mechanisms needs to be adopted for linking the archived digital news articles.

The study proposes an effective term-based approach for linking digital news articles in DNSA. The approach is empirically analyzed, and the results of the proposed approach are compared to get conclusive arguments.

## 2 Background

Enormous information is available on the web for users, including a variety of products and options in the form of books, restaurants, hotels, research articles, movies, news articles, etc. Recommender systems help users to focus down information to manageable sets. Broadly, two approaches have been devised, Collaborative Filtering approach, based on similar users having similar interest or same demographics and Content-based approach, which is based on features of the item to be recommended [2, 13].

It is observed that the news articles available can be very huge and recommendation systems can help to recommend relevant news to news readers by filtering news articles based on predefined criteria. There are two approaches that can be used to link news articles together either by collaborative filtering or content based approach. Collaborative filtering method presents many challenges as it relies on the similarity in opinions and demographics of the users [6]. It becomes more complicated with dynamic nature of the users and news articles themselves. Users prefer to find recent news in online news environment, which is difficult to learn user's preferences that lead an accurate model based on the items they previously read [1, 7, 11]. User preferences and interest changes over time, depends on the current events and popularity of the news articles themselves [12]. Typically, users are not willing to click, to recommend news articles

during browsing news on the specific topic [16]. Content-based approach recommends new items to the user based upon the similarity value being computed between the descriptions or features of items selected previously. Content based approach can run through its own problems like determine similarity between news articles that represents different topics and the way user's choice effect by some potentially hidden factors [10]. All these studies focused on run time similarity between recent articles. The subsequent section describes the linkage of preserved news articles in the DNSA.

## 2.1 Content-Based Similarity Approach

Collaborative filtering approach suffers through a number of challenges because of many reasons. In contrast, content-based filtering approach can be used to adopt features of the item, i.e. news article to be recommended or link articles together by some predefined criteria of similarity among the news stories by extracting features from news articles to avoid third party dependency, i.e. user feedback, etc. Almost all the studies about recommending news articles focuses on run time similarity and recommendation because these platforms have dealt with few numbers of recently published articles. The subsequent section describes the linkage of preserved news articles in the DNSA.

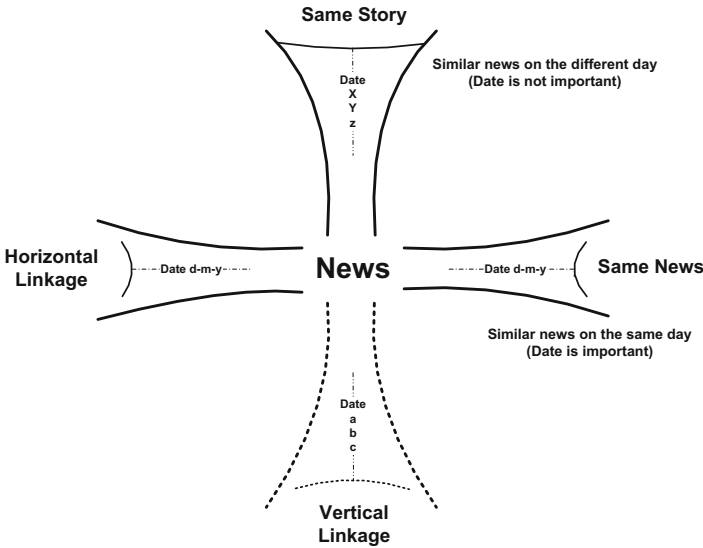
## 3 Linking Digital News Stories in DNSA

The Digital News Stories Archive (DNSA) has passed more than six months of the trial period. The number of extraction has passed greater than 100 times at different intervals in which initially three locals leading English online newspapers (Dawn News, The Tribune and The News) being considered for preservation and the number increased to ten, that include seven local English online newspapers and three local news television networks, which provide online English news to the news readers. The archive (created locally) currently preserving more than one thousand news after removing duplicate URLs and news in each extraction.

The news readers read about an event or an issue from various sources in order to get a broader perspective and diverse viewpoints that help to better understand the world around. Moreover, consulting various sources helps in authenticating the information by comparing similar news from multiple news sources. The DNSA has news articles from multiple sources, needs to create a mechanism that helps the reader to read a set of relevant news stories about an event or issue. The DNSA needs an efficient mechanism to link the digital stories and recommend to the readers. This linkage will lead the reader to browse through the huge collection easily. Without a suitable and efficient linking mechanism for relevant news, the newspaper or online newspaper archive is nothing more than a data collection.

A link can be created by two means, namely vertical linkage and horizontal linkage as presented in Fig. 1.

1. **Vertical Linkage:** The link is created between the news stories which represent the “same story” regardless of the time frame based on all the news preserved in the DNSA. Same story can be interpreted as the follow-up news about an issue from the same source or from different sources. For example, news stories about the court proceedings of a particular case. Same story linkage leads to follow-up similarity between news for more days since the start of proceedings.
2. **Horizontal Linkage:** The link is created between the news stories which represent the “same news” based on all the news preserved in the DNSA in a specific time period i.e. day, week, month. Same news can be interpreted as the similar news reported about an issue from the same source or from different sources. For example, Court response to the arguments about a particular case. Same news linkage makes the identification of news about the same topic easy.

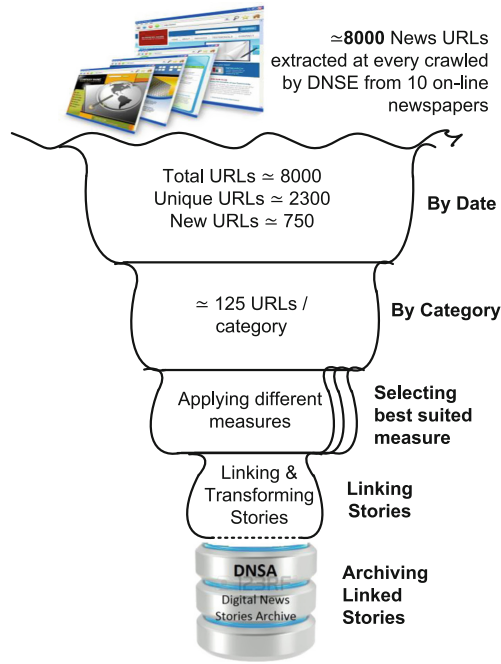


**Fig. 1.** Linking digital news stories in DNSA

In this study, the focus is to create links between relevant stories based on similar terms used in the news articles, comprehensively discussed subsequently.

The news articles are linked in the DNSA during the preservation process in the form of implicit metadata to easily relate news stories based on horizontal linkage. It is observed that two news stories about an event may contain similar terms and published online on the same day, which show the importance of terms used and encourage utilizing these terms for linking digital news stories during preservation in the Digital News Stories Archives.

To speed up the linking process of digital news stories for preservation, a tiered approach is adopted as presented in Fig. 2.



**Fig. 2.** Layered approach to reduce workload during linking digital stories for preservation

The main purpose of the tiered approach is to reduce the linking workload and improve performance. Currently, the DNSA archives 10 online newspapers and more than 8000 URLs are extracted at every crawl by DNSE. In the first tier, the news URLs are filtered by date, so that removing duplicate URLs and news already archived from previous crawls. In the second tier, the news is filtered by categorizing the news articles into six known categories which reduce the news article collection sufficiently for linking. Third tier can be divided into multiple tiers based on different measures developed for linking digital news stories for preservation. The selection of measure depends upon the performance of that measure in a specific category, e.g. a measure can perform better on a specific category, such as “Opinion”. In the last tier, the news stories are linked and transformed to the digital news stories archive.

## 4 Term-Based Approach

To find out the similarity between news in the DNSA, needs to process the terms used in the news articles. The news article contains the different type of terms,

e.g. nouns, verbs, adverbs, etc. In academic journals, the nouns are considered to be the main key phrases [14] but other terms like verbs and adverbs, etc., also play a vital role in representing the news articles [4]. Therefore, introduced a Common Ratio Measure for Stories (CRMS) based on the similar terms used in the English news articles except stop words for linking digital news stories during preservation. The CRMS Algorithm 1 pseudo-code is given as follows;

---

**Algorithm 1.** CRMS Algorithm

---

- 1: **News article pre-processing**
  - 2: Filtering non-news contents and extracts the news article from the news webpage
  - 3: **Compute Term Frequencies**
  - 4: **repeat**
  - 5:     Tokenize news articles using StanfordCoreNLP
  - 6:     Remove stop words
  - 7:     Calculate term frequencies of each term in the news articles
  - 8: **until** Both the news articles are processed
  - 9: **Compute CT, UT and TT**
  - 10: Compute CT (Common Terms) Count
  - 11: Select all common terms in both the news articles with frequencies
  - 12:  $CT = (tf_1 + tf_2)W_1 + (tf_1 + tf_2)W_2 + \dots + (tf_1 + tf_2)W_n$
  - 13:  $CT = \sum_{i=1}^n (tf_1 + tf_2)W_i$
  - 14:     ▷ Where,  $W_i$  is the common term or word in both the selected news articles,  $tf_1$  term frequency of word  $W$  in one news,  $tf_2$  is term frequency of word  $W$  in second news article and  $n$  is the total number of common terms in both the news.
  - 15: Compute UT (Uncommon Terms) Count
  - 16: Select all uncommon terms in both the news articles with frequencies
  - 17:  $UT = (tf_1 \vee tf_2)W_1 + (tf_1 \vee tf_2)W_2 + \dots + (tf_1 \vee tf_2)W_m$
  - 18:  $UT = \sum_{j=1}^m (tf_1 \vee tf_2)W_j$
  - 19:     ▷ Where  $m$  is the total number of uncommon terms in both the news
  - 20: Compute TT (Total Terms) Count
  - 21:  $TT = CT + UT = \sum_{i=1}^n (tf_1 + tf_2)W_i + \sum_{j=1}^m (tf_1 \vee tf_2)W_j$
  - 22:     ▷ The total terms in both the news articles
  - 23: **Compute Common Ratios**
  - 24: Four common ratios can be used as similarity measure between news articles
  - 25:  $CRMS = CT/TT$
  - 26:  $CRMS = CT/UT$
  - 27:  $CRMS = UT-CT$
  - 28:  $CRMS = UT/TT$
- 

*CT/TT*: The value varies between 1 ( $CT = TT$ ) and 0 ( $CT = 0$  i.e.  $UT = TT$ ). 1 means the news are exact copies of each other and 0 represents that the stories are completely different. Closer the value  $CT/TT$  to 1 leads to more similarity and

closer the value to 0 decrease similarity. The interpretation of measure  $UT/TT$  is exactly opposite to the measure  $CT/TT$ . The value  $CT$  needs to be normalized by the value  $TT$  because the high value of  $CT$  only does not show the accurate similarity.

For example, if  $CT$  of two news A and B is 150 and  $CT$  of news A and C is 100 do not show that the A and B is more similar than A and C until normalized. If  $TT$  count for A & B is 400 and for A & C is 200 then by  $CT/TT$  are 0.375 and 0.5 respectively, shows that news articles A & C is more similar than A & B.

$CT/UT$ : The value varies between 0 ( $CT = 0$ ) and  $TT$ ,  $UT > 0$ , greater the value similar the news articles and vice versa.

$UT - CT$ : The minimum possible value is  $-TT$  and maximum value is  $+TT$ . Minimum the value leads to high similarity of the news articles and maximum the value leads to dissimilarity.

## 5 Evaluating CRMS

The size of the DNSA is growing very fast as on average seven hundred new articles are added every day. The number may rise even further when more news sources are added. The size of the DNSA makes it an ideal choice for evaluating the CRMS. Various news sets were created for evaluating the significance of CRMS.

### 5.1 News Sets

The following four sets were defined which were created multiple times for evaluating the CRMS.

- **Set 1:** A set containing three news articles - two similar and one dissimilar. Each article in this set is manually chosen from a different newspaper.
- **Set 2:** A set containing ten news articles - three closely related, one partially similar and six dissimilar. The articles in this set are manually chosen from three different newspapers.
- **Set 3:** A set containing thirty news articles - the news articles are categorized based on six topics. Table 1 presents the details of the stories in each category. The articles in this set are manually chosen from nine different newspapers.
- **Set 4:** A set containing 215 news articles - the articles are grouped into two categories. First, contains 52 articles related to sports which were automatically extracted from the DNSA using the category information. Second contains 163 automatically extracted news articles related to other topics including politics, business, law and order situation, crimes and cases in courts. The articles in this set are chosen randomly from three different newspapers.

**Table 1.** News articles distribution in 30 news articles dataset

S.No	Topic	No of news
Topic 1	Disruptive passenger in PIA at Heathrow London	6
Topic 2	Trump travel ban	5
Topic 3	CPEC	5
Topic 4	Nurses protest in Karachi	4
Topic 5	Earthquake in Baluchistan	5
Topic 6	LoC ceasefire violation	5

## 5.2 Evaluation Methods

Two evaluation methods, namely system centric (automatic) and user centric (human judgment) were used for assessing the accuracy and effectiveness of the designed similarity measure for linking news stories for preservation.

### 1. Pre-experiment: Human Judgment of Similarity

In this section the similarity level by human judgment is defined. The similarity between the news articles has defined empirically by two means, i.e. online news readers and by an expert. The online news readers included graduate and undergraduate students as well as faculty members who were randomly selected to rank the similarity among news articles based on human judgment. An individual is selected as an expert who knows comprehensively about the DNSE and the process of digital news preservation in DNSA.

Three sets of news articles were created and used to evaluate the proposed approach. Each set contained ten news articles selected from different online news publishers. The collection contained news articles that were selected by reading headings from the same genre i.e. sports, opinions, entertainment. The detail of each news set is presented in Sect. 5.1.

To define similarity among news articles, the participants were asked to define similarity of one news article with the rest of the news articles in the collection using five-point Likert scale and a representation of numerical measure from 1 to 10. The basic reason to consider the numerical scale to easily differentiate similar or dissimilar news i.e. if news B & C are marked similar to that of news A in Likert scale; the same may be marked as 10 and 9 respectively at the numeric scale that shows the news B is more similar to that of news C. The following Table 2 show the numerical equivalent values of likert scale.

**Table 2.** Numeric scale to corresponding likert scale

Likert scale	Similar	Partially similar	Unsure	Partially dissimilar	Dissimilar
Numeric scale	10–9	8–7	6–5	4–3	2–1



2. Automatic Evaluation

Precision and recall were calculated and compared with the proposed similarity measure i.e. CRMS. *Precision* is the ratio of the number of relevant items retrieved to the total number of irrelevant and relevant items retrieved. *Recall* is the ratio of the number of relevant items retrieved to the total number of relevant items in a collection. Let A is No of relevant items retrieved, B is No of irrelevant items retrieved and C is No of relevant items not retrieved, the set-based measures are given in Table 3.

Table 3. Set-based measures

	Relevant	Not relevant
Retrieved	A	B
Not retrieved	C	D

$$\text{Precision} = A / A + B(Rel\&Ret / Retrieved)$$
  
It is usually expressed in percentage i.e.  
$$\text{Precision} = (A / A + B) * 100$$
  
$$\text{Recall} = A / A + C(Rel\&Ret / Relevant)$$
  
It is usually expressed in percentage i.e.  
$$\text{Recall} = (A / A + C) * 100$$

5.3 Results

The proposed term based approach i.e. CRMS, for similarity between news articles is analyzed on different sets of news articles. The overview of datasets is summarized in the Table 4.

Table 4. Overview: Datasets of news articles used for evaluation

S.No	News articles			Similarity observed			
	No of news articles/set	No of sets	No of newspapers	During selection	By news reader	By expert	Proposed measures
1	3	3	3	Yes	No	No	Yes
2	10	3	3	Yes	Yes	Yes	Yes
3	30	1	9	Yes	No	No	Yes
4	215	1	3	No	No	No	Yes

For each set of news articles, similarity is computed and measured by two means, i.e. empirically (User based) and Automatic (using CRMS algorithm). The Tables 5, 6 and 7 shows the summary of the values computed for evaluation and the similarity between news articles are compared.

An experiment was performed using a set of articles collected following the definition of set three (defined in Sect. 5.1). The performance of CRMS for the set is presented in the Table 8.

**Table 5.** Similarity comparison (Likert scale) with CRMSs for Set 1

	News Readers		Expert		Common Ratio Measure for Stories (CRMS)							
News1	News	Mean	News	Value	CT/TT	Value	UT/TT	Value	CT/UT	Value	UT-CT	Value
ns1	ns3	4.7	ns3	5	ns3	0.542	ns3	0.458	ns3	1.181	ns3	-29
ns1	ns8	4.3	ns8	5	ns8	0.409	ns8	0.591	ns8	0.691	ns8	81
ns1	ns5	3.6	ns5	4	ns7	0.333	ns7	0.667	ns7	0.498	ns7	149
ns1	ns10	2.9	ns10	4	ns10	0.262	ns10	0.738	ns10	0.355	ns10	151
ns1	ns7	2.7	ns7	4	ns5	0.222	ns5	0.778	ns5	0.285	ns9	166
ns1	ns9	2.6	ns9	4	ns9	0.165	ns9	0.835	ns9	0.198	ns4	177
ns1	ns4	2.4	ns4	4	ns6	0.161	ns6	0.839	ns6	0.192	ns5	181
ns1	ns6	1.4	ns6	1	ns4	0.147	ns4	0.853	ns4	0.173	ns2	201
ns1	ns2	1.3	ns2	1	ns2	0.135	ns2	0.865	ns2	0.155	ns6	282

**Table 6.** Similarity comparison (Likert scale) with CRMSs for Set 2

	News Readers		Expert		Common Ratio Measure for Stories (CRMS)							
News1	News	Mean	News	Value	CT/TT	Value	UT/TT	Value	CT/UT	Value	UT-CT	Value
ns1	ns5	4.6	ns5	5	ns5	0.281	ns5	0.719	ns5	0.39	ns5	97
ns1	ns9	4.3	ns9	4	ns2	0.158	ns2	0.842	ns2	0.187	ns4	116
ns1	ns10	3.4	ns10	4	ns8	0.154	ns8	0.846	ns8	0.182	ns3	119
ns1	ns7	1.7	ns7	1	ns7	0.148	ns7	0.852	ns7	0.174	ns2	152
ns1	ns6	1.6	ns6	1	ns10	0.141	ns10	0.859	ns10	0.164	ns6	152
ns1	ns2	1.4	ns2	1	ns9	0.128	ns9	0.872	ns9	0.147	ns8	157
ns1	ns3	1.4	ns3	1	ns6	0.116	ns6	0.884	ns6	0.131	ns9	215
ns1	ns8	1.4	ns8	1	ns3	0.066	ns3	0.934	ns3	0.07	ns7	223
ns1	ns4	1.2	ns4	1	ns4	0.032	ns4	0.968	ns4	0.033	ns10	229

**Table 7.** Similarity comparison (Likert scale) with CRMSs for Set 3

	News Readers		Expert		Common Ratio Measure for Stories (CRMS)							
News1	News	Mean	News	Value	CT/TT	Value	UT/TT	Value	CT/UT	Value	UT-CT	Value
ns1	ns3	4.4	ns3	5	ns3	0.688	ns3	0.312	ns3	2.203	ns3	-273
ns1	ns7	4.1	ns2	5	ns7	0.41	ns7	0.59	ns7	0.694	ns7	89
ns1	ns2	3.7	ns5	5	ns5	0.407	ns5	0.593	ns5	0.686	ns5	91
ns1	ns4	3.7	ns7	4	ns2	0.405	ns2	0.595	ns2	0.68	ns4	118
ns1	ns5	3.6	ns4	4	ns4	0.372	ns4	0.628	ns4	0.593	ns2	129
ns1	ns8	3.5	ns8	4	ns8	0.308	ns8	0.692	ns8	0.445	ns8	177
ns1	ns9	3.4	ns9	4	ns6	0.256	ns6	0.744	ns6	0.345	ns6	251
ns1	ns10	2.9	ns10	4	ns10	0.256	ns10	0.744	ns10	0.345	ns9	278
ns1	ns6	2.4	ns6	4	ns9	0.175	ns9	0.825	ns9	0.212	ns10	287

The results of the evaluation show that the CRMS gives reliable results. Therefore, it is useful to use the CRMS for linking digital news stories. An archive containing preserved news stories which are linked using the proposed measure i.e. CRMS will support accessibility of related news articles.

To measure the precision and recall of the proposed similarity measure i.e. CRMS, the experiment is performed on a set of articles collected following the definition of Set three (defined in Sect. 5.1). The similarity is observed by an expert (manually) during the selection of news articles for the experiments as presented in Table 4. The performance of CRMS for the dataset is presented in the Table 8.

**Table 8.** Precision and recall for CRMS

S.No	Topic	Precision	Recall
Topic 1	Disruptive passenger in PIA at Heathrow London	100%	100%
Topic 2	Trump travel ban	80%	100%
Topic 3	CPEC	60%	75%
Topic 4	Nurses protest in Karachi	60%	100%
Topic 5	Earthquake in Baluchistan	80%	100%
Topic 6	LoC ceasefire violation	80%	100%

The results of the evaluation show that the CRMS gives reliable results. Therefore, it is useful to use the CRMS for linking digital news stories. An archive containing preserved news stories which are linked will support accessibility of related news articles. The digital news stories normalized, linked and preserved ensure accessibility of related news articles.

## 6 Conclusions and Future Work

The proposed measure i.e. CRMS is simple to calculate as it involves just a few steps. However, the proposed algorithm may not produce accurate results when used for two articles having very different lengths i.e. compare a two sentence story with a full page story. However, the findings show that the similarity results based on the CRMS are able to capture the reality when used for articles not very different in length.

Currently, work is going on to extend the linkage of stories to news published in Urdu language. Moreover, work is in progress to develop tools for exploiting the linkage created among stories during the preservation process for search and retrieval tasks.

## References

1. Agarwal, D., Chen, B.-C., Elango, P., Wang, X.: Personalized click shaping through lagrangian duality for online recommendation. In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 485–494. ACM (2012)
2. Athalye, S.: Recommendation system for news reader (2013)
3. Burda, D., Teuteberg, F.: Sustaining accessibility of information through digital preservation: a literature review. *J. Inf. Sci.* **39**(4), 442–458 (2013)
4. Chun, D.: On indexing of key words. *Acta Editologica* **16**(2), 105–106 (2004)
5. da Silva, J.R., Ribeiro, C., Lopes, J.C.: A data curation experiment at U. Porto using DSpace (2011)
6. Doychev, D., Lawlor, A., Rafter, R., Smyth, B.: An analysis of recommender algorithms for online news. In: CLEF (Working Notes), pp. 825–836. Citeseer (2014)
7. Fortuna, B., Fortuna, C., Mladeníć, D.: Real-time news recommender system. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010. LNCS (LNAI), vol. 6323, pp. 583–586. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15939-8\\_38](https://doi.org/10.1007/978-3-642-15939-8_38)
8. Khan, M., Ur Rahman, A.: Digital news story preservation framework. In: Proceedings of Digital Libraries: Providing Quality Information: 17th International Conference on Asia-Pacific Digital Libraries, ICADL 2015, Seoul, Korea, 9–12 December 2015, vol. 9469, p. 350. Springer (2015). <https://doi.org/10.1007/978-3-319-27974-9>
9. Khan, M., Ur Rahman, A., Daud Awan, M., Alam, S.M.: Normalizing digital news-stories for preservation. In: 2016 Eleventh International Conference on Digital Information Management (ICDIM), pp. 85–90. IEEE (2016)
10. Li, L., Li, T.: News recommendation via hypergraph learning: encapsulation of user behavior and news content. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, pp. 305–314. ACM (2013)
11. Li, L., Wang, D.-D., Zhu, S.-Z., Li, T.: Personalized news recommendation: a review and an experimental investigation. *J. Comput. Sci. Technol.* **26**(5), 754–766 (2011)
12. Li, L., Zheng, L., Yang, F., Li, T.: Modeling and broadening temporal user interest in personalized news recommendation. *Expert Syst. Appl.* **41**(7), 3168–3177 (2014)
13. Melville, P., Sindhvani, V.: Recommender systems. In: Sammut, C., Webb, G.I. (eds.) *Encyclopedia of Machine Learning*, pp. 829–838. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-0-387-30164-8\\_705](https://doi.org/10.1007/978-0-387-30164-8_705)
14. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Proceedings of the Fifth ACM Conference on Recommender Systems, pp. 157–164. ACM (2011)
15. Ur Rahman, A.: Data warehouses in the path from databases to archives. Ph.D. thesis, Faculty of Engineering, University of Porto, July 2013
16. Said, A., Bellogín, A., Lin, J., de Vries, A.: Do recommendations matter?: news recommendation in real life. In: Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 237–240. ACM (2014)