Creating Event-Centric Collections from Web Archives

Elena Demidova and Thomas Risse

Abstract Web archives are an essential information source for research on historical events. However, the large scale and heterogeneity of web archives make it difficult for researchers to access relevant event-specific materials. In this chapter, we discuss methods for creating event-centric collections from large-scale web archives. These methods are manifold and may require manual curation, adopt search or deploy focused crawling. In this chapter, we focus on the crawl-based methods that identify relevant documents in and across web archives and include link networks as context in the resulting collections.

1 Introduction

Web archive materials are of high relevance for research on historical events, such as the Olympic Games, political elections and traumatic incidents with high societal impact. Research on the past events is of interest for many disciplines, including digital humanities, history, law and journalism (Risse et al. 2014b). Given an event of interest and a large-scale web archive, researchers have difficulty in identifying relevant archived content accurately and efficiently. In this chapter, we discuss methods that automatically create focused, event-centric collections for studies on historical events from large-scale web archives.

E. Demidova (✉)
Data Science & Intelligent Systems, Computer Science Institute, University of Bonn, Bonn, Germany

L3S Research Center, Leibniz University of Hannover, Hannover, Germany
e-mail: demidova@L3S.de

T. Risse
University Library J. C. Senckenberg, Goethe University Frankfurt, Frankfurt am Main, Germany
e-mail: t.risse@ub.uni-frankfurt.de

D. Gomes et al. (eds.), *The Past Web*,
https://doi.org/10.1007/978-3-030-63291-5_6

Services for the creation of event-centric or topical collections may require manual curation (e.g. the Archive-It service[1]), as well as adopt search and focused crawling (e.g. Farag et al. 2018; Gossen et al. 2015a). These methods, initially developed for the live Web, are increasingly adopted to the creation of event-centric collections derived from web archives.

Manual curation typically means that the user specifies the web addresses (Uniform Resource Identifiers [URIs]) of interest to be included in the collection. A variation of this method can include search, where the URIs retrieved by a search engine are verified by the user and included in the collection. Manual curation can be very precise, in particular in cases where the intended collection covers only a few relevant URIs known in advance. It also allows for the preservation of the link context of a document. However, manual curation becomes unfeasible as the number of relevant URIs increases and as the temporal gap grows between the event and the time of collection creation. Nevertheless, manual curation can provide valuable input for automatic collection-creation methods. This includes, for example, the seed URIs for crawl-based approaches, which are discussed later in this chapter.

Search-based collection-creation methods can make use of full-text search over the archived documents (see chapter "Full-Text and URL Search" for more details), or of keyword search in the lightweight indexes constructed over the URIs that have been proposed in Souza et al. (2015). In these cases, a collection corresponds to a set of search results retrieved from a web archive and consists of snapshots of web resources. While search-based approaches can appear intuitive, as they correspond to the way users look for information on the live Web, they have several shortcomings. First, in the search-based methods, users' intent is typically expressed as a keyword query. However, keyword queries possess low expressivity and do not adequately capture the complex semantics of event-centric collections. Second, search retrieves document snapshots in isolation and does not provide contextual information for the linked documents. Finally, search-based methods require a priori full-text indexing of the archived resources, which is still not supported by many web archives.

In contrast, the crawl-based methods recently proposed by Gossen et al. (2017, 2020) and extended by Klein et al. (2018) adapt focused crawling, initially designed for the live Web, to web archives. These methods address the shortcomings discussed earlier and provide the following advantages: first, crawl-based techniques allow for more expressive collection specification, which reflects the topical and temporal properties of the events; second, event-centric collections created through crawl-based methods preserve the document linking structure within the archived web graph, thus providing context; and third, crawl-based methods do not require any full-text indexes. Recently, crawl-based methods have been applied to create event-centric collections from a single web archive (Gossen et al. 2017, 2020) as well as to create collections combining relevant documents from several web archives (Klein et al. 2018).

[1] http://archive-it.org.

This chapter summarises recent works on the creation of event-centric collections from large-scale web archives, with a focus on crawl-based methods. In the following sections, we discuss the definition and specification of event-centric collections, present collection-creation methods in more detail and discuss open challenges.

2 Definitions of Event-Centric Collections

The aim of creating event-centric collections from web archives is to gather documents relevant to a given historical event. Such collections can contain—as a result—a set of identifiers of the archived documents (URIs) or a set of snapshots of the archived webpages.

When defining event-centric collections, we distinguish between the event and its representation on the Web. The philosopher Kim defined an event as a happening that has a finite duration and changes object's properties or relations (Kim 1976). Westermann and Jain proposed a more formal model that defined an event as a tuple of aspects (Westermann and Jain 2007). These aspects are informational (e.g. entities), structural (e.g. sub-events), spatial (e.g. places, regions, coordinates), temporal (e.g. physical time), causal (chain of causing events) and experiential (e.g. documenting media, sensor data). Farag et al. adopted this model for event-centric collection creation by crawling webpages on the live Web (Farag et al. 2018). There are many other variations of event definition in the literature. In this chapter, we adopt Kim's view.

Events can exhibit different granularity and significance. The target user group for the creation of event-centric collections from web archives are researchers in digital humanities, history, law, journalism and other disciplines. These researchers are typically concerned with more significant events that are reflected on the Web and are of some interest for society. Therefore, events considered in this chapter are happenings of historical relevance, such as the Olympic Games, political elections and traumatic incidents with high societal impact. In contrast, fine-grained events like Mrs Smith moved to Berlin will rarely be reflected on the Web, unless Mrs Smith was a prominent person. A focus on events of societal significance raises the chance that at a later point in time an overview, a summary page or an article will appear, e.g. on Wikipedia, that summarises the event and provides pointers to other relevant materials. These pointers can be used in the crawl specification as a starting point to collect related documents (i.e. as an initial seed list, as discussed later in Sect. 4).

According to Kim's event definition, time and objects changing their properties are essential event characteristics. In the case of significant events, the Web typically adequately reflects such crucial event properties. Shortly after the beginning of an event, the number of mentions of related objects rapidly increases on the Web and decreases subsequently, depending on the level of public interest. Primarily, this is the case for unexpected events such as natural disasters and terrorist attacks.

Gossen et al. observed that for planned, and in particular for recurring events (sport competitions, political elections), relevant documents often appear ahead of the actual start of an event, during the event lead time, and are still published after the completion of the event, during the cool-down time (Gossen et al. 2017, 2020).

Another aspect of collection creation is the user perspective on collection usage. Holzmann and Risse identified three views: user-, data- and graph-centric (Holzmann and Risse 2017). In the user-centric view, users access the archived pages by specifying a URI and time, or by using full-text search. The focus is on the full presentation of a single page for close reading. In the data-centric view, webpages are considered as single resources and treated as raw data, such as text or images. Collections following the data-centric view consist of mostly isolated pages that are not necessarily connected. Finally, in the graph-centric view, single pages or websites, consisting of multiple resources, are considered as nodes in a graph, without taking their contents into account. In such a graph, edges between the nodes represent the links among the pages. Chapter "A Holistic View on Web Archives" discusses these views in more detail.

The user perspective identified by Holzmann and Risse (2017) can provide an essential basis for the collection-creation strategy. For example, a special kind of collection that follows the user-centric view is the storytelling proposed by AlNoamany et al. (2017). This method summarises smaller events and turns them into a story. The stories are built by sampling the most relevant pages from the collection and ordering them by time. The users can access the results via a storytelling social media interface like Adobe Livefyre (previously Storify). Search-based methods can appropriately support the data-centric view. For example, Nanni et al. (2017) followed the data-centric view and selected specific isolated documents. In contrast, the graph-centric view requires a crawl-based method, which preserves the links in the archived web graph. Gossen et al. (2017, 2020) proposed a crawl-based approach that facilitates the graph-centric view. Klein et al. extended this work to accumulate data from several distributed web archives in the resulting collection (Klein et al. 2018).

3 Event-Centric Collection Specification

To create an event-centric collection from a web archive, the user needs to define the desired collection characteristics and the parameters required to configure the specific collection-creation method.

Relevant characteristics of the target collection can include the topical relevance and the temporal scope of the documents as well as the domains of interest. Methods for collection creation can be search or crawl based or involve manual curation. Parameters that delimit a crawl-based collection can include seed URIs of relevant pages to start the crawling process, as well as further restrictions like the number of links followed from the seed URIs.

To facilitate a canonical specification of these characteristics and parameters, Gossen et al. introduced a Collection Specification (Gossen et al. 2017, 2020). This specification, initially developed for crawl-based collection creation, provides a rather generic view, whereas other collection-creation methods typically adopt a subset of the characteristics and parameters specified in the Collection Specification. An exemplary Collection Specification for the event *Snowden leaks* from Gossen et al. (2020) is presented in Example 1.

Example: Collection Specification 1 (Snowden Leaks) *This collection can be described as:*

- *Topical Scope:*
  - *reference documents: wikipedia:Edward_Snowden*
  - *keywords: nsa, edward snowden, ...*
- *Temporal Scope:*
  - *time span of the event: [2013-06-01, 2013-06-30];*
  - *duration of the lead time* $T_l = 0$*;*
  - *duration of the cool-down time* $T_r = 1\ year$*.*

The Collection Specification describes the topical and temporal scope of relevant documents to be included in the event-centric collection. Like the approaches for search and focused crawling on the live Web proposed in Gossen et al. (2015b), a Collection Specification can indicate the intended topical scope through keywords and named entities. Furthermore, examples of relevant documents can narrow down the collection scope. Recent approaches to event-centric collection creation typically adopt Wikipedia articles on the event of interest as examples (Gossen et al. 2017, 2020; Klein et al. 2018), whereas in general any relevant documents can be named.

The temporal scope is the time interval during which relevant documents appear on the live Web or are stored in a web archive. In the case of events, the relevant time frame depends on the time at which the event took place as well as on the nature of the event (see Table 1). For example, regularly recurring events (e.g. sport competitions), planned special events (e.g. concerts) or unexpected events (e.g. natural disasters) indicate specific temporal reporting patterns (Gossen et al. 2020). Therefore, existing approaches include in the Collection Specification the lead time,

Table 1 Examples of temporal event characteristics

Event	Type	Duration	Lead time	Cool-down time
Olympic Games	Recurring	2 weeks	Weeks	Days
Federal election	Recurring	1 day	Months	Weeks
Fukushima accident	Non-recurring	1 week	–	Months
Snowden leaks	Non-recurring	1 day	–	Years

i.e. the time ahead of the event, and the cool-down time after the completion of the event during which relevant documents appear on the Web.

Furthermore, a Collection Specification for crawl-based collection creation can include specific instructions and parameters for the crawling procedure. For example, the Collection Specification may include seed pages for the crawler and crawl depth (Gossen et al. 2017, 2020; Klein et al. 2018). The seed URIs may constitute the entire collection, like the manually curated services of Archive-It, or serve as starting points for web archive crawling. Web archive crawling may adopt existing link collections as seed URIs. For example, recent approaches utilise links extracted from Wikipedia event articles as seeds for the focused archive crawler (Gossen et al. 2017, 2020; Klein et al. 2018). In future work, possible extensions might include the use of search results as seed URIs, like (Bicho and Gomes 2016), thus combining search-based and crawl-based collection-creation methods.

The choice of the relevant version of the Wikipedia article from which to obtain an appropriate event representation varies across the approaches. For example, Gossen et al. selected the most recent version of the Wikipedia article, assuming it to be more complete (Gossen et al. 2017, 2020). In contrast, Klein et al. adopted temporal heuristics based on the article's editing behaviour to pick the version reflecting the original event representation (Klein et al. 2018).

Further relevant parameters are the intended collection size and the relevance threshold for the included documents. Termination conditions may include a time limit for the collection creation. Gossen et al. developed a user interface to support the specification of crawl-based event-centric collections for the live Web (Gossen et al. 2015b). Such interfaces may be adapted to define the Collection Specification for web archives in the future.

4 Methods to Create Event-Centric Collections

In this section, we briefly discuss methods to create event-centric collections from web archives, including manual curation, search and crawling.

One rather straightforward method to create collections is to search the content of the web archive or its metadata. Existing web archives, such as the UK Web Archive,[2] the Arquivo.pt[3] archive or the Internet Archive,[4] increasingly support full-text search, or metadata search as described in chapter "Full-Text and URL Search". Search results can directly provide a data-centric view of the archived data (see Sect. 2), but the surrounding context of a page might be missing. Nanni et al. (2017) expanded the collections to include related aspects found in knowledge bases like DBpedia (Lehmann et al. 2015). Here, specialised event-centric knowledge

[2] https://www.webarchive.org.uk/ukwa/.

[3] https://arquivo.pt/?l=en.

[4] https://archive.org/.

graphs, such as EventKG (Gottschalk and Demidova 2018, 2019), can potentially provide an important source of information regarding events and their temporal relations. One particular case of metadata search is semantic URI search, which adopts named entities mentioned in the URIs of the archived documents to identify relevant documents efficiently while relying on a lightweight index structure, as opposed to full-text indexing (Souza et al. 2015).

Strategies that aim at a graph-centric or user-centric view need to follow the link structure in the web archive, like regular web crawlers. Whereas Gossen et al. created crawl-based collections from a single web archive (Gossen et al. 2020, 2017), Klein et al. (2018) used the Memento protocol (Bornand et al. 2016) to facilitate collection creation from 22 web archives. Web archive crawlers have five main phases: seed selection, content fetching, content assessment, link extraction and prioritisation. The crawlers typically fetch and analyse the pages one by one. Alternatively, a crawler may fetch all pages of the seed list at once and apply all the subsequent steps in the same way (e.g. Nanni et al. (2017)).

Seed selection may be performed manually, by utilising a reference source like a Wikipedia article with manually curated links to relevant documents, or by using a search engine. All methods have their pros and cons. Fully manual curation of the seed list provides collection curators with the highest level of control. However, this strategy will fail to include relevant seeds not known to the curators. A prominent example of manually curated web archive collections is Archive-It,[5] even though the seed list of this service is used for live web crawling. Manual curation approaches are in principle feasible for web archive collections, although not widely adopted. A semi-automatic approach is to draw on a Wikipedia article of the event of interest and utilise the links that it contains as seeds (Gossen et al. 2017; Klein et al. 2018). However, Wikipedia-based seed selection is only possible for significant events regularly curated in Wikipedia. Web search engines typically index major parts of the Web, such that their search results probably include the relevant seeds for essential events. However, the factors influencing their result ranking are not transparent. Nonetheless, web search engines can provide useful hints for a seed list. As web search engines operate on today's Web, past content is often not visible to them. Search engines operating directly on web archive content can circumvent this problem.

The content-fetching phase in web archives is similar to content fetching on the Web. Given a URI, the crawler selects the corresponding archived page. Nevertheless, the ephemeral nature of webpages leads to substantial differences while crawling these media. The live Web contains only one, the most recent version of a page. In contrast, web archives contain different versions of the same page, as a crawler might visit a page at multiple points in time. Therefore, the evolution of a page, e.g. a live blog about an event, is documented at different times. The question that arises is which version should be selected. However, this cannot be answered in general, as it depends on the anticipated usage of the collection. For

[5] https://archive-it.org/.

example, to analyse the evolution of perception of an event, it might be interesting to observe the evolution of individual page content over time. In contrast, to obtain a comprehensive event overview, the newest version of a page might be a better choice as it typically contains the most recent updates.

Content assessment is a crucial phase in the collection-creation process. In this phase, given the Collection Specification, the crawler will assess the relevance of a page. The result of the assessment is a relevance score, used to decide if a page will become part of the collection. The threshold for this decision depends on collection intent. A high threshold leads to a precise page selection, with the risk of leaving out relevant pages. With a lower threshold, the number of relevant pages increases. However, a lower threshold will also lead to the inclusion of further non-relevant pages. Identification of a suitable threshold value typically requires some experimentation to tune the selection and crawl parameters.

There are different approaches to content assessment. Gossen et al. and Klein et al. used a vector representation of an event (Gossen et al. 2017; Klein et al. 2018). Gossen et al. created an event vector from the Wikipedia event article and used this vector as a reference to judge the relevance of the crawled pages (Gossen et al. 2017). They created a candidate vector for each crawled page, similar to the creation of the event vector. The crawler then calculated the relevance score as the cosine similarity between the event vector and the candidate vector. The impact of specific terms mentioned in the Collection Specification on relevance estimation can be increased by additional factors such as the inverse document frequency (IDF). IDF values can be obtained, for example, from the Google NGram datasets.[6]

To our knowledge, machine learning has not yet been applied in the context of crawl-based collection creation from web archives. In contrast, on the live Web, several approaches have been developed for topic-centric crawling of webpages. Such focused crawlers use text classification to learn a model to estimate the relevance of an unseen page (Chakrabarti et al. 1999; Pant and Srinivasan 2005). Reinforcement learning is used in Rennie and McCallum (1999) and Singh et al. (2012) by modelling the crawling process as a Markov decision process. Menczer and Monge combined genetic programming and reinforcement learning to estimate page relevance (Menczer and Monge 1999). Adoption of machine learning to crawl-based collection creation in web archives is a promising direction for future research.

Besides content relevance, temporal relevance also plays an important role. Event-related documents are published not only during the event time interval but also before and after. Consequently, we need to estimate the relevance of a document based on the Collection Specification and a time point associated with the web document (e.g. the creation, last modification or capture date). Like Kanhabua and Nørvåg (2011), Gossen et al. assumed that the relevance of documents decreases rapidly as the distance from the event increases and therefore defined a temporal relevance function based on the exponential decay function (Gossen et al. 2020).

[6] https://books.google.com/ngrams.

The document time point can be estimated using the *document focus time* discussed in Jatowt et al. (2013).

The link extraction phase extracts all links included in the page. These are the links to related pages, pictures, layout information and scripts. If the collection adopts a user-centric or a graph-centric view, this collection should include layout information of the pages. The crawler adds the extracted links to the crawl queue and can also apply prioritisation, e.g. based on the page score or through further content analysis.

Demidova et al. proposed methods for semantic enrichment of event-centric collections using knowledge bases (Demidova et al. 2014). They applied these methods to event-centric collections created from the live Web in the ARCOMEM project (Risse et al. 2014a). Further development of these methods to enrich event-centric collections created from web archives, in particular through the adoption of event-centric knowledge graphs such as EventKG (Gottschalk and Demidova 2018, 2019), is a promising direction for future research.

5 Discussion and Open Research Directions

In this chapter, we discussed methods aimed at the creation of event-centric collections from web archives. One important consideration when selecting a specific approach for collection creation is the purpose of the collection from the user's perspective, where we differentiate among the user-, data- and graph-centric views proposed by Holzmann and Risse (2017). This view determines the parameters of the creation method and the characteristics of the selected pages, in particular concerning their interlinking and relevant context in the archived web graph.

Furthermore, the collection-creation methods may be classified as manual curation, search-based or crawl-based. In this chapter, we paid particular attention to crawl-based techniques. We observe that existing processes for the creation of collections from the live Web can successfully be adapted to web archives. In addition, aspects such as temporal relevance, to select the suitable version of a document, and seed selection, to reach documents that may not exist on the live Web, should be addressed.

Finally, we observed that state-of-the-art methods in the area of machine learning have not yet been widely adopted in the context of collection creation from web archives. Here, we see significant potential for future research. Furthermore, recently proposed event-centric knowledge bases, like EventKG (Gottschalk and Demidova 2018, 2019), or entity-centric knowledge bases, like DBpedia (Lehmann et al. 2015) and Wikidata (Vrandecic and Krötzsch 2014), provide a rich source of event- and entity-centric semantic information to complement archived materials and to facilitate further advances in event-centric collection creation.

References

AlNoamany Y, Weigle MC, Nelson ML (2017) Generating stories from archived collections. In: Proceedings of the 2017 ACM conference on web science, ACM, WebSci '17, pp 309–318

Bicho D, Gomes D (2016) Preserving websites of research & development projects. In: Proceedings of the 13th international conference on digital preservation, iPRES 2016

Bornand NJ, Balakireva L, de Sompel HV (2016) Routing memento requests using binary classifiers. CoRR abs/1606.09136

Chakrabarti S, van den Berg M, Dom B (1999) Focused crawling: a new approach to topic-specific Web resource discovery. Comput Netw 31(11–16):1623–1640

Demidova E, Barbieri N, Dietze S, Funk A, Holzmann H, Maynard D, Papailiou N, Peters W, Risse T, Spiliotopoulos D (2014) Analysing and enriching focused semantic web archives for parliament applications. Fut Intern 6(3):433–456

Farag MMG, Lee S, Fox EA (2018) Focused crawler for events. Int J Digit Lib 19(1):3–19

Gossen G, Demidova E, Risse T (2015a) iCrawl: improving the freshness of web collections by integrating social web and focused web crawling. In: Proceedings of the 15th ACM/IEEE-CE joint conference on digital libraries. ACM, New York, pp 75–84

Gossen G, Demidova E, Risse T (2015b) The iCrawl wizard - supporting interactive focused crawl specification. In: Proceedings of the 37th European conference on IR research, ECIR 2015. Lecture Notes in Computer Science, vol 9022, pp 797–800

Gossen G, Demidova E, Risse T (2017) Extracting event-centric document collections from large-scale web archives. In: Proceedings of the 21st international conference on Theory and Practice of Digital Libraries, TPDL 2017, pp 116–127

Gossen G, Demidova E, Risse T (2020) Towards extracting event-centric collections from Web archives. Int J Digit Lib 21(1):31–45

Gottschalk S, Demidova E (2018) EventKG: a multilingual event-centric temporal knowledge graph. In: Proceedings of the ESWC 2018, pp 272–287

Gottschalk S, Demidova E (2019) EventKG - the hub of event knowledge on the web - and biographical timeline generation. Seman Web 10(6):1039–1070

Holzmann H, Risse T (2017) Accessing web archives from different perspectives with potential synergies. In: Researchers, practitioners and their use of the archived web (London, 2017)

Jatowt A, Yeung CA, Tanaka K (2013) Estimating document focus time. In: Proceedings of the 22nd ACM international conference on information and knowledge management, CIKM'13. ACM, New York, pp 2273–2278

Kanhabua N, Nørvåg K (2011) A comparison of time-aware ranking methods. In: Proceeding of the 34th international ACM SIGIR conference on research and development in information retrieval, SIGIR 2011. ACM, New York, pp 1257–1258

Kim J (1976) Events as property exemplifications. Springer Netherlands, Dordrecht, pp 159–177

Klein M, Balakireva L, de Sompel HV (2018) Focused crawl of web archives to build event collections. In: Proceedings of the 10th ACM conference on web science, WebSci 2018. ACM, New York, pp 333–342

Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, Hellmann S, Morsey M, van Kleef P, Auer S, Bizer C (2015) DBpedia - a large-scale, multilingual knowledge base extracted from Wikipedia. Seman. Web 6(2):167–195

Menczer F, Monge AE (1999) Scalable web search by adaptive online agents: an InfoSpiders case study. In: Klusch M (ed) Intelligent information agents. Springer, Berlin, Heidelberg, pp 323–347. https://doi.org/10.1007/978-3-642-60018-0_17

Nanni F, Ponzetto SP, Dietz L (2017) Building entity-centric event collections. In: Proceedings of the 2017 ACM/IEEE joint conference on digital libraries, JCDL 2017. IEEE Computer Society, Washington, pp 199–208

Pant G, Srinivasan P (2005) Learning to crawl: comparing classification schemes. ACM Trans Inf Syst 23(4):430–462

Rennie J, McCallum A (1999) Using reinforcement learning to spider the web efficiently. In: Proceedings of the sixteenth international conference on machine learning (ICML 1999). Morgan Kaufmann, Burlington, pp 335–343

Risse T, Demidova E, Dietze S, Peters W, Papailiou N, Doka K, Stavrakas Y, Plachouras V, Senellart P, Carpentier F, Mantrach A, Cautis B, Siehndel P, Spiliotopoulos D (2014a) The ARCOMEM architecture for social- and semantic-driven web archiving. Fut Intern 6(4):688–716

Risse T, Demidova E, Gossen G (2014b) What do you want to collect from the web? In: Proceedings of the building web observatories workshop (BWOW) 2014

Singh N, Sandhawalia H, Monet N, Poirier H, Coursimault JM (2012) Large scale URL-based classification using online incremental learning. In: Proceedings of the 2012 11th international conference on machine learning and applications, ICMLA '12, vol 02. IEEE Computer Society, Washington, pp 402–409

Souza T, Demidova E, Risse T, Holzmann H, Gossen G, Szymanski J (2015) Semantic URL analytics to support efficient annotation of large scale web archives. In: First COST Action IC1302 International KEYSTONE conference IKC, 2015. Lecture notes in computer science, vol 9398. Springer, New York, pp 153–166

Vrandecic D, Krötzsch M (2014) Wikidata: a free collaborative knowledgebase. Commun ACM 57(10):78–85

Westermann U, Jain R (2007) Toward a common event model for multimedia applications. IEEE MultiMedia 14(1):19–29