Int J Digit Libr (2018) 19:77–93
https://doi.org/10.1007/s00799-016-0200-8

Avoiding spoilers: wiki time travel with Sheldon Cooper

Shawn M. Jones[1] · **Michael L. Nelson**[2] · **Herbert Van de Sompel**[1]

Received: 1 March 2016 / Revised: 21 November 2016 / Accepted: 25 November 2016 / Published online: 21 December 2016

Abstract A variety of fan-based wikis about episodic fiction (e.g., television shows, novels, movies) exist on the World Wide Web. These wikis provide a wealth of information about complex stories, but if fans are behind in their viewing they run the risk of encountering "spoilers"—information that gives away key plot points before the intended time of the show's writers. Because the wiki history is indexed by revisions, finding specific dates can be tedious, especially for pages with hundreds or thousands of edits. A wiki's history interface does not permit browsing across historic pages without visiting current ones, thus revealing spoilers in the current page. Enterprising fans can resort to web archives and navigate there across wiki pages that were live prior to a specific episode date. In this paper, we explore the use of Memento with the Internet Archive as a means of avoiding spoilers in fan wikis. We conduct two experiments: one to determine the probability of encountering a spoiler when using Memento with the Internet Archive for a given wiki page, and a second to determine which date prior to an episode to choose when trying to avoid spoilers for that specific episode. Our results indicate that the Internet Archive is not safe for avoiding spoilers, and therefore we highlight the inherent capability of fan wikis to address the spoiler problem internally using existing, off-the-shelf technology. We use the spoiler use case to define and analyze different ways of discovering the best past version of a resource to avoid spoilers. We propose Memento as a structural solution to the problem, distinguishing it from prior content-based solutions to the spoiler problem. This research promotes the idea that content management systems can benefit from exposing their version information in the standardized Memento way used by other archives. We support the idea that there are use cases for which specific prior versions of web resources are invaluable.

Keywords Digital preservation · HTTP · Resource versioning · Web archiving · Wikis · Spoilers

✉ Shawn M. Jones
smjones@lanl.gov

Michael L. Nelson
mln@cs.odu.edu

Herbert Van de Sompel
herbertv@lanl.gov

[1] Los Alamos National Laboratory, Los Alamos, NM, USA

[2] Old Dominion University, Norfolk, VA, USA

1 Introduction

From *How I Met Your Mother* to *Game of Thrones*, fans have created wikis covering their favorite episodic fiction. The plots of television shows have become increasingly complex where each episode is often part of a larger story rather than just a self-contained narrative. This has led to a need for resources that have documented characters, plots, and other items from these fictional worlds. For a fan base, the wiki becomes that focal point for continued discussion and documentation of the fictional milieu. The fan wiki phenomena has been researched, with the first study focusing on the *Lostpedia* wiki, which covered the TV series *Lost* (2004–2010) [20].

There has also been a rise in the availability of streaming services such as *Netflix* and media recording devices such as *Tivo*. While hardcore fans will do everything in their power to view a new episode when it first airs, these technologies allow fans to catch up afterwards, viewing at a later time. The availability of this technology has created a problem of abundance by making much more content available than ever

before. Because this content can be viewed at the time of the fan's choosing, fans will defer viewing one show because they are already viewing another, pushing their time of consumption further from the air date. In some cases, shows appear in different countries at different times (e.g., BBC's *Downton Abbey* and *Sherlock* air in the USA several months later than in the UK). These are all reasons why a fan's time of viewing may be deferred from the air date.

For a given television show's aficionado, using the fan wiki entails the danger of ruining the experience of watching the episode, because those fans that have viewed the initial airing will already have updated the pages. As a result, the fans that have not caught up encounter *spoilers*.

Spoilers are defined as chunks of information for which a fan wants to control time and place of consumption, preferring to consume them in the order intended by the fictional work. If this information is consumed in the wrong order, the enjoyment of the work is diminished [13]. The spoiler problem has been reported in popular media for years by a variety of sources including *CNN* [11] and *The New York Times* [23]. *Wired* emphasizes how fans of many television series prefer to view their content as the creators intended, rather than having it ruined by parts of the Web [12].

Steiner, van Hooland, and Summers [24] found that, generally, new events lead to updates to wiki pages. Applying this observation specifically to a fan wiki suggests that a new episode leads to edits of the wiki pages that reflect the evolution the plot underwent during the episode. Hence, it is safe to assume that revisions to a fan wiki page made after the first airing of an episode are very likely to contain spoilers pertaining to that episode. Fans are aware of this problem and try to avoid these spoilers [22]. In fact, fans are often outraged when other media unexpectedly provides spoilers.[1]

In essence, these fans are looking for revisions of the wiki pages published prior to the date that the episode was aired. Resources exist to help them determine that date.[2] As it comes to prior revisions, they exist in the fan wiki itself because it typically stores all page revisions. However, accessing prior revisions in a wiki is cumbersome because it entails consulting the page history, seen in Fig. 1, scrolling to find a date prior to airing, and then consulting a prior revision of the page. More importantly, the page history approach has to be repeated for every page of interest because links in a prior revision of a page lead to the current revision of the linked page and hence…to spoilers [7], as seen in Fig. 2. Prior revisions of the wiki pages also exist in web archives as a result of ongoing independent web crawling activities.[3] By far the largest web archive is the Internet Archive. It started

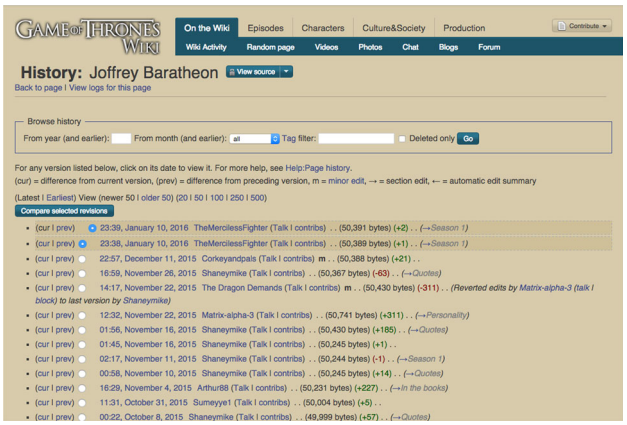

Fig. 1 Browsing the history of a wiki page in order to avoid spoilers is cumbersome

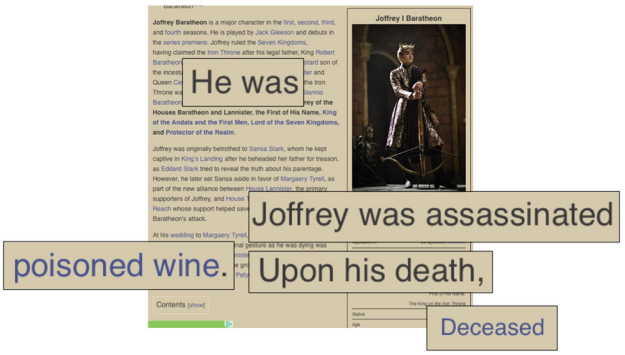

Fig. 2 Fans can only access the history of a wiki page from the current revision of the page, which often contains unavoidable spoilers, in this case the death of the character Joffrey Baratheon

collecting web pages in 1996 [25], meanwhile stores about 440 billion web pages, and responds to about 82 million consultation requests on a daily basis [5].

Memento (RFC 7089) [27] is a web protocol that allows users to find the version of a web resource at a specific moment in time. The Memento protocol is an extension to the Hypertext Transfer Protocol (HTTP) that allows a user to specify a web site and a datetime and then redirects the user to the best past version of that page based on the datetime given by the user. It provides seamless access to all systems where previous versions of web pages are stored, from web archives to content management systems. An in depth discussion of Memento is outside of the scope of this paper; for more information consult [27–29]. With solutions such as the Wayback Machine,[4] users slide from one datetime to another while continuing to click links to archived pages, but Memento holds a fan's datetime fixed, ensuring that all links lead to pages based on their desired date and time [1].

Research exists on using web archives, like the Internet Archive, to retrieve lost content [5]. Examples include

[1] http://comicbook.com/blog/2013/02/08/the-big-bang-theory-comes-under-fire-for-walking-dead-spoiler/.

[2] http://epguides.com.

[3] https://archive.org/details/web&tab=collection.

[4] http://wayback.archive.org.

deleted political party content[5] and lost legal decisions [31]. Others have used the Internet Archive to identify which items in social networks are missing but still retrievable [4,21].

In this paper, we explore whether it is wise for fans to use Memento with the Internet Archive in order to avoid spoilers. As the intent of Memento is to find the version of a web resource at a specific moment in time, a fan should be able to use Memento to select a datetime prior to the missed episode of their favorite show and arrive at a version of a fan wiki page without spoilers.

This work makes several contributions. We use the spoiler use case to define and analyze different ways of discovering the best past version of a resource to avoid spoilers. We propose Memento as a structural solution to the problem, distinguishing it from prior content-based solutions to the spoiler problem [6,10,13]. This research promotes the idea that content management systems, such as wikis, are archives in their own right and hence can benefit from exposing their version information in the standardized Memento way used by other archives. We support the idea that there are use cases for which specific prior versions of web resources are invaluable. Even though we study the problems in terms of spoilers, our solution of using Memento can also satisfy a number of other use cases, such as delaying the consumption of sports scores and studying news within a historical context.

2 Related work

In 2011, Leavitt and Christenfeld conducted a study where 819 participants took part in three experiments [18]. They were given stories to read. For each story, the researcher created a spoiler paragraph describing the story and revealing the "outcome in a way that seemed inadvertent". If a subject had already read a story, their data for that story was excluded from the experiment. Each version of each story was rated on a 10-point scale, where 10 was considered *best*. Unexpectedly, slightly more participants preferred spoiled stories over unspoiled stories. The study also indicated that readers are unable to compare spoiled and unspoiled experiences and thus those who preferred spoiled stories may just prefer spoilers in general.

Schirra, Sun, and Bently conducted a study of live-tweeting while the television show *Downton Abbey* was airing [22]. Live-tweeting is a process whereby those watching a television show episode discuss the show on a social media web site, such as Twitter, while the episode is airing. This study consisted of a sample of 2,234 participants who live-tweeted during the highly anticipated third season premier and beyond. The intention of the study was to determine how long fans continued to engage in live tweeting after the first episode. They discovered a complex social process with its own evolving rules and customs. Semi-structured interviews were conducted among some of the participants. Some of the United Kingdom live-tweeters would hold off revealing spoilers, but still live-tweet during the American air dates so that they could vicariously share in the story reveals and plot information as their American friends experienced it. Others would concoct methods to communicate major plot twists, such as using ambiguous pronouns, without spoiling the story for their friends. Because the broadcast can experience propagation and transmission delays, some live-tweeters had the show spoiled by others because their friends' experience differed by a matter of a minute or less, resulting in tweets that arrived to the tweeter before they actually got to experience the topic of the tweet. Some live-tweeters would avoid social media altogether, finding that their experience could still be spoiled by others. In one case, a live-tweeter stopped watching the show once another Twitter user spoiled it for them. *Downton Abbey* represents an example of a global problem because it airs in the United Kingdom months prior to the United States. As seen in Fig. 3, maintainers of the *Downton Abbey wiki* post a spoiler notice until Americans are able to catch up.

Fig. 3 Wiki page editors admit that fans want to avoid spoilers through the use of spoiler notices; in this case fans may be from a country where the show has not yet aired

This is consistent with a study conducted by Johns, also using interviews in a small group of participants who also engaged in *two screen viewing*, a more generic name for live-tweeting [14]. In this study Johns discovered that those who used digital video recording (DVR) devices, such as the TiVO, would avoid social media until they had watched their show. Also, some would eschew DVRs because they wanted to participate in live-tweeting. This kind of frustration with spoilers indicates a social problem that refutes the results of Leavitt and Christenfeld.

[5] http://ws-dl.blogspot.com/2013/11/2013-11-21-conservative-party-speeches.html.

Leaver described "The Tyranny of Digital Distance", further emphasizing the issue of television shows airing in one country months before another [17]. Leaver discusses the same issue experienced by the American *Downton Abbey* fans, but this time with the television show *Battlestar Galactica*. Leaver mentions how *Battlestar Galactica* aired in the United States 6 months or longer prior to airing in Australia. He argues that the Internet provides near instantaneous communications between fans of a television show, but the broadcast and distribution networks for television content do not engage in a simultaneous release of content, resulting in fans experiencing spoilers because other fans live in a different time zone, or in a country where legal issues are delaying the release of content. He even mentions that using news sources, such as *Google News*,[6] can result in spoilers for those who live in a different country than the one creating the content. He also refutes the argument, put forth by Leavitt and Christenfeld, that spoilers do not affect the enjoyment of fiction, by mentioning that plot leaks for the *Harry Potter* novels were disastrous for fans, resulting in public outcry [8].

Because of the phenomenon of spoilers in social media, Boyd-Graber, Glasgow, and Zajac conducted an evaluation of machine learning approaches to find spoilers in social media posts [6]. They used classifiers on multiple sources to determine which posts should be blocked. They determined that spoilers are identified by transitive words, such as "kill", that affect the outcome of a plot because they link characters to each other. They also mention that spoilers refer to events "*later* than the viewer's knowledge of the current work", suggesting that any machine learning technique used for avoiding spoilers in social media must be smarter than just blocking all posts about a particular topic [10,13]. Their classifiers were trained by crowdsourcing and acquiring data from the Internet Movie Database,[7] TV Tropes,[8] and Episode Guides[9] online resources. By utilizing these additional sources, they were able to use machine learning techniques to identify spoilers better than their predecessors, who relied primarily on term matching and small data sets. Inspired by this work were software packages that block spoilers from a fan's social media feed, such as Spoiler Shield[10] and the Netflix Spoiler Foiler [9].

Mittell conducted an extensive study of *Lostpedia* [20], the fan wiki site for the television show *Lost*. He highlights how robust and accurate a fan wiki becomes as more fans improve upon its content, even indicating how pages can contain spoilers. He discussed the social nature of wikis and indicates how post-episode traffic exists, much like the findings of Steiner, van Hooland, and Summers with real-world events inspiring Wikipedia edits [24].

Almedia, Mozafari, and Cho produced one of the first studies of the behavior of contributors to Wikipedia [3]. The authors discovered that there are distinct groups of Wikipedia contributors. They suggest that, as the number of articles increases, the contributors' attention is split among more and more content, resulting in a larger number of contributors creating revisions of existing pages than creating new ones. This informs our use, in this paper, of the number of edits as a surrogate to the popularity of a page.

We highlight this academic work to demonstrate that there is appetite for analyzing and solving the spoiler phenomenon for web content. The prior solutions suggest a content-based approach to solving the spoiler problem for fans, requiring training data for a variety of television shows that are known in advance. These solutions become more difficult to use at scale, as more and more content needs to be analyzed. We describe the use of a structural-based approach instead, which will work without such training and can apply to any topic for which one wants to avoid spoilers, including sports scores. Our approach scales and is less prone to error, because it is purely based on timestamps and uses the standard protocols of the web.

We are exploring the spoiler avoidance concept as related to fan wikis, not social media. Also, our intent is not to block wiki pages, but rather to make use of prior version of those pages that avoid spoilers and yet are still useful resources for fans.

Of particular interest to advertisers are the cases where viewers abandon shows, or online content, due to spoilers. For this reason, there is an actual financial benefit to content producers to remedy these problems [26]. We believe that wiki administrators and other content producers can help their fans by using Memento to ameliorate this problem.

3 Revisions, observations, mementos

Past versions of web pages are referred to as **mementos**. The Memento Protocol (RFC 7089) provides datetime negotiation—a variant on content negotiation—to allow access to a memento for the original resource that was the live web version at or around the preferred datetime. The user selects a datetime for their request, and the archive then redirects them to the best memento for that preferred datetime. The selected datetime in the request is expressed in a special-purpose HTTP protocol request header from RFC 7089, named accept-datetime, in the same form as other HTTP content negotiation headers (e.g., accept, accept-language, accept-language). In light of this, we will use the term **accept-datetime** to indicate the desired datetime of the fan.

In order to avoid spoilers for an episode, a fan requires:

[6] http://news.google.com.

[7] http://imdb.com.

[8] http://tvtropes.org.

[9] http://epguides.com.

[10] http://www.spoilershield.com.

Fig. 4 Television episodes exist as a series of events on a timeline

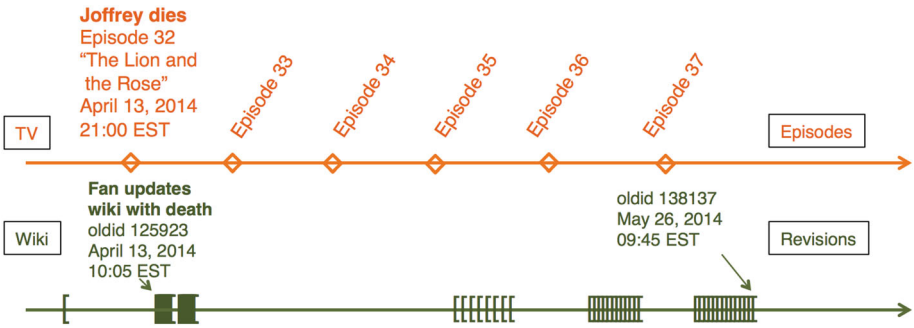

Fig. 5 Wiki revisions exist as a series of events on a different timeline, but spoilers in episodes, in this case Joffrey's death, serve as a catalyst for the creation of new wiki revisions by fans

– The datetime of the original airing of an episode.
– Access to a version of a wiki page prior to that datetime.
– Sticking to the same datetime when navigating from page to page by clicking links.

Versions of a wiki page exist in the fan wiki itself as well as in web archives such as the internet archive. However, there are significant differences:

– **Observations** Robots sent out by a web archive crawl the web and ingest encountered resources, including wiki pages. Robots revisit wiki pages every now and then. But pages can change in between visits and only those versions of the page that were visited are stored in the archive. Hence, web archives contain distinct observations, but are not aware of the evolution of a page between these observations [1,2].
– **Revisions** A wiki typically stores all revisions for a page (*oldid* pages in MediaWiki terminology). It records the page's and the wiki's history.

Both observations and revisions, as past versions of web resources, are mementos. The date that a web page is first active is stored in a special-purpose HTTP header named **memento-datetime**. In the case of the observation, the memento-datetime is when it was observed. In the case of the revision, the memento-datetime is when that specific revision was created.

We can visualize the temporal relationship between these concepts as a series of timelines. First there exist real-world events, in our case, television episodes, represented as a timeline in Fig. 4. In this figure, we see multiple episodes of *Game of Thrones*. In Episode 32, the character of Joffrey Baratheon dies, which is a spoiler to fans of the television show who have not yet watched this episode.

Wiki revisions exist on a different timeline. As seen in Fig. 5, the events within episodes on the top timeline serve as a catalyst for the creation of revisions of wiki pages shown on the timeline below it. Within hours of Episode 32, we see that a fan has updated the character's wiki page with information about Joffrey's death. We also see 52 additional edits after this episode and prior to the next episode.

While the first edit at revision 125923 in Fig. 6 merely marks the character as *deceased*, subsequent separate edits by other fans:

– change the verb tense to the past tense in sentences referring to the character
– indicate how and where the character has died—he was poisoned at his wedding
– update the titles of other characters as a result of his death—e.g., the Queen now becomes the Queen Regent
– change the references to which episodes and seasons he appears—e.g., he was listed as a fourth season character but now is listed as appearing in only the first two episodes of the fourth season.

This is consistent with the pattern mentioned by Steiner et al. [24].

Independent of episodes and revisions is a third timeline, shown on the bottom in Fig. 6, indicating observations by a web archive. The web archive crawls the web in search of content, and is not informed by the wiki of new revisions. This bottom timeline shows an observation that was created 87 revisions after Joffrey's death was recorded in the wiki. Another observation, not shown due to the scale, exists two weeks prior.

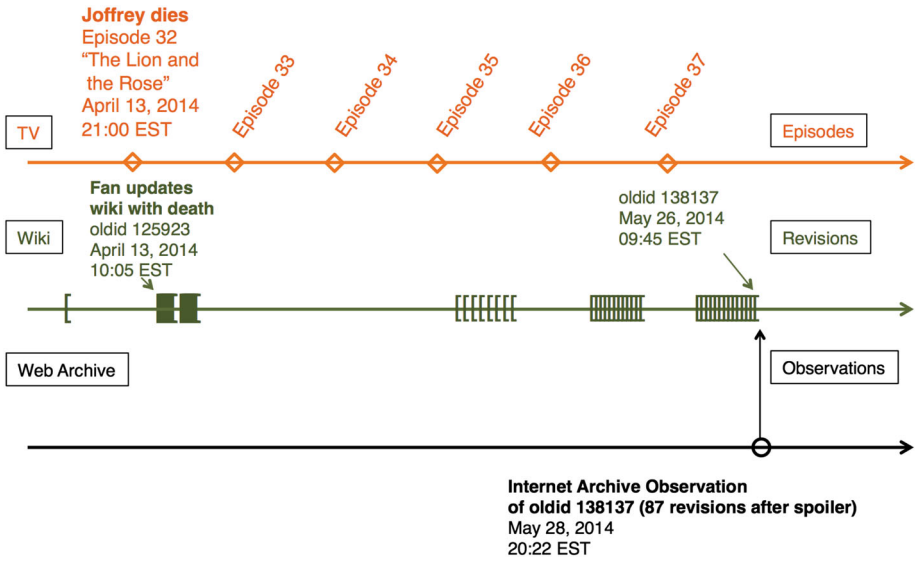

Fig. 6 The web archive makes observations of specific revisions of a given wiki page; in this case, this observation is of a revision created long after Joffrey's death

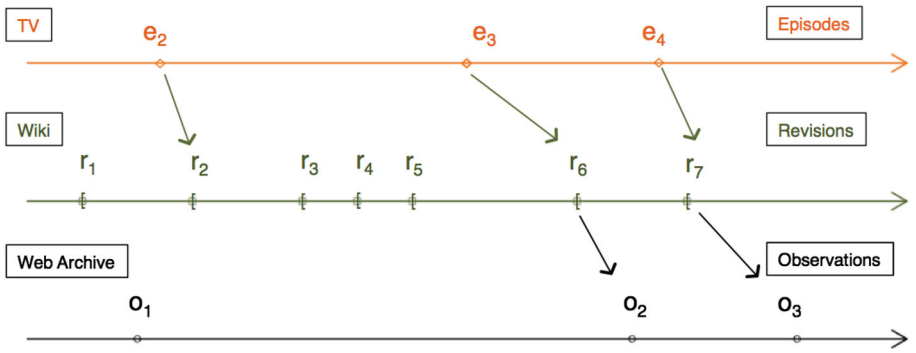

Fig. 7 The three timeline model used in this paper: television episodes inspire fan wiki revisions which are recorded as web archive observations; in this example, e_3 inspires wiki revision r_6, which is recorded as observation o_2

The relationship between these three timelines serves as a basis for understanding the nature of the spoiler problem and the relationships between real-world events, in this case television episodes, wiki revisions created from those episodes, and the eventual web archive observations of those revisions (Fig. 7).

3.1 Observations: approximate time travel

Archives typically miss revisions, while wikis do not. As a result, accessing a revision of a page for a specific datetime (i.e., a datetime prior to that of the first airing of an episode) is a different challenge for a wiki than for a web archive.

Because a web archive records observations, and is not aware of the evolution of a page between these observations, it has no means to unambiguously determine which revision was operational at any given datetime. Here, we use the term **operational** to mean available to any visiting user of the site at that time. As a matter of fact, it may not even have an observation of that revision. As a result, a web archive provides an approximate response based on the observations it stores and uses the **mindist** (short for *minimum distance*) heuristic, seen in Fig. 8, to select an observation. Using mindist, the observation that is temporally closest to the requested datetime will be selected, irrespective of whether that observation is before or after the requested datetime. For example, in Fig. 8, choosing a datetime just prior to e_3 will redirect the fan forward to observation o_2, but choosing one just prior to r_4 will direct the fan backward to observation o_1, missing information from revisions r_2 through r_4.

It is the combination of these two characteristics of web archives—recording observations and using the mindist heuristic—that can cause spoilers in web archives even when accessing them with a date prior to that of the initial airing of the episode. For example, in Fig. 8, observation o_2 contains spoilers with respect to e_3 because it is an observation of r_6, which was created past the time of airing e_3. If a fan chooses an accept-datetime just prior to e_3, observation o_2 is presented to the user, because it is temporally closest to e_3.

Mindist is a logical heuristic for web archives, because they typically fail to observe many revisions of a page. Therefore, a user would want the closest observation available to their accept-datetime because the dates of capture may be wildly distant from one another.

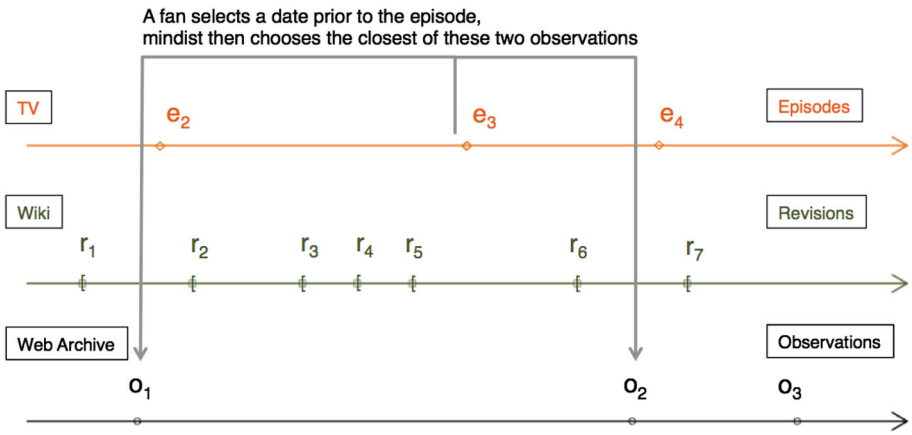

Fig. 8 The mindist heuristic used by web archives to direct a fan to the best observation; if the fan chooses a datetime just prior to episode e_3, mindist must choose between observations o_1 and o_2 and will return observation o_2 because that one is closest to the fan's datetime

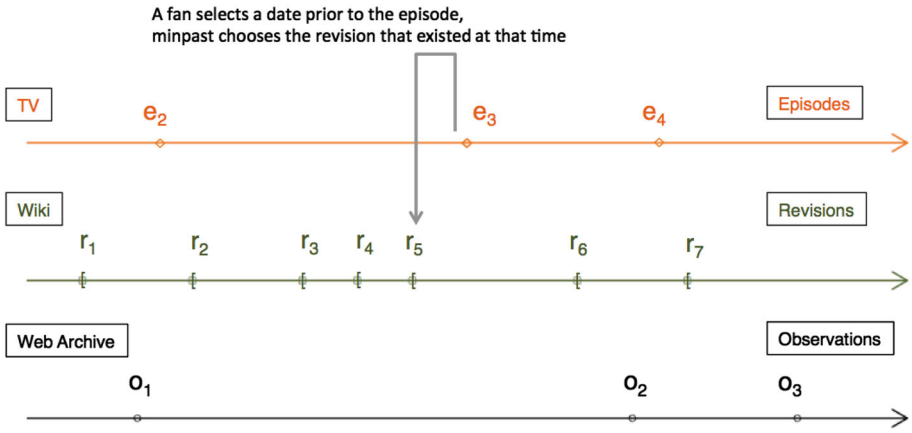

Fig. 9 Example of the minpast heuristic, showing a fan choosing an accept-datetime just prior to episode e_3 and being directed to the most current revision r_5 at that time, avoiding spoilers for episode e_3

Consider a real-world example involving the university department web site http://www.cs.odu.edu. One observation was captured by the Internet Archive on October 6, 2008[11] and the next observation was captured on January 21, 2009.[12] Now, let the typical user choose an accept-datetime of December 20, 2008. Because the Internet Archive has few observations to choose from during this time period, the January 21, 2009 observation is best in this case, because it is most likely to represent the general time period for which the user was interested.

3.2 Revisions: certain time travel

Because a wiki records history [19], it is possible to unambiguously determine which revision of a page was operational at a given datetime. For example, in Fig. 8, we see all revisions r_1 through r_7 and we can see that at the time episode e_2 was aired, the active revision was r_1. Therefore, **minpast** (short for *minimum distance in the past*), is the logical choice when implementing the Memento Protocol for a wiki. Because wikis are dense archives that contain every revision, minpast, shown in Fig. 9, can more successfully help Memento achieve its goal of providing the version—in this case a wiki revision—of a web resource at a specific moment in time. Because its holdings are sparse in comparison to wikis, the Internet Archive does not have enough observations to make minpast a viable default.

4 The probability of avoiding spoilers in the internet archive

Consider a theoretical fan who wants to avoid spoilers in their favorite fan wiki and does so by consulting the Internet Archive instead. This fan is particularly aggressive, when looking for archived wiki pages, and chooses as values for Memento's accept-datetime every second between the first episode and the latest episode. Using the memento-datetimes of the observations to which the fan is redirected by means of the Internet Archive's mindist heuristic, we can calculate the probability of encountering a spoiler for a given page.

To determine when our fan encounters a spoiler for a given wiki page, we need:

[11] https://web.archive.org/web/20081006041026/ http://www.cs.odu.edu/.

[12] https://web.archive.org/web/20090121122424/ http://cs.odu.edu/.

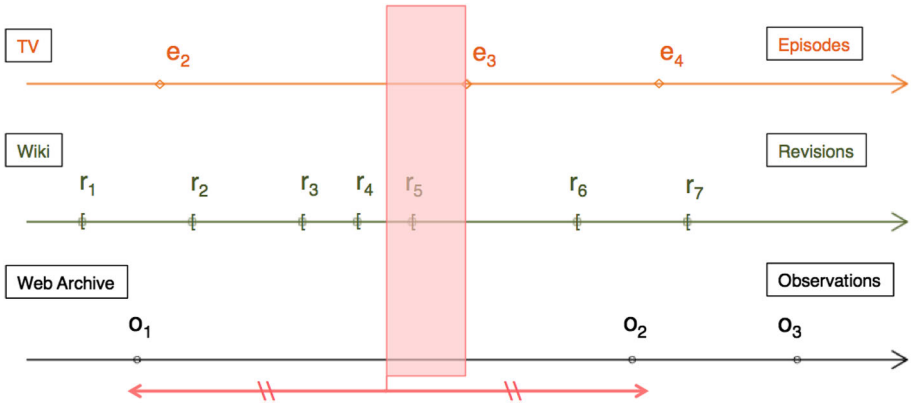

Fig. 10 Graphical representation of a spoiler area (shown in *red*), for episode e_3; if a fan selects an accept-datetime anywhere in this region, they will be directed to observation o_2, which comes from revision r_6 created after the episode; because it is caused by mindist, the spoiler area extends from the midpoint between two observations to the datetime of the episode (color figure online)

– The datetimes of the original airing of each episode.
– The creation datetime of every revision of the wiki page.
– The memento-datetimes of every observation of the wiki page in the Internet Archive.

Once we have that information, for each page we perform the following:

1. Choose the datetime of the first episode as the accept-datetime.
2. Record the memento-datetime of the observation that is returned according to the mindist heuristic used by the Internet Archive.
3. If the memento-datetime of the observation is greater than the creation datetime of the first revision after the episode, then record it as a spoiler; in essence

$$\text{Spoiler}(t) = \begin{cases} 1 & \text{if } o_k \geq r_{\text{first after episode } e_i} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where t is the accept-datetime under consideration, o_k is the kth observation, and $r_{\text{first after episode } e_i}$ is the first revision for the wiki page created after episode e_i.
4. Increment accept-datetime by 1 s, and repeat for all seconds until the latest episode of the series

The areas in the timeline where a fan encounters a spoiler, even though they chose an accept-datetime prior the episode, are defined as **spoiler areas**, visualized in Fig. 10. Because of the nature of mindist, the spoiler area is bounded by the midpoint between two observations and the episode for which they are trying to avoid spoilers.

Probability is defined as the number of times something can occur divided by the total number of outcomes [30]. Let c be the number of seconds between the first and latest episodes in a series, then the probability of encountering a spoiler for a given wiki page is shown by equation (2).

$$Pr(\text{spoiler}) = \frac{\sum_{t=1}^{n} \text{Spoiler}(t)}{c} \quad (2)$$

Once we have determined the probability of encountering a spoiler for a resource within the Internet Archive, we can use it to compare that resource to others. In this way we can determine how safe a given URI is for fans who want to avoid spoilers using the Internet Archive.

We selected 15 fan wikis, shown in Table 1, based on weekly television shows, for our experiment. Each television show selected has had at least two seasons and a currently active wiki. As mentioned before, *Lost* was also included because its wiki, *Lostpedia*, has undergone academic study, and is the oldest and largest fan wiki under consideration.

We then calculated spoiler areas and probabilities for all pages in each wiki by collecting, for each page, the creation datetime of every revision and the memento-datetime of observations of the page in the Internet Archive. We collected episode air dates from epguides.com. All revision and observation data were gathered September and October of 2014.

Table 2 shows a breakdown of the data reduction for our experiment. Out of the 40,617 wiki pages processed for this experiment, we discovered that many of them were wiki redirects. Redirects are used to deal with articles that can be referred to by multiple names. Sometimes fans may not know the real name of an introduced character until much later, and will use a redirect from the old name to the new. Sometimes fans will create pages for a topic not knowing that one already exists, leaving future fans to create a redirect now that they know that a new page title was desired. Because of the number of redirects that contained only a single revision and only a single observation, we removed the redirects from consideration for calculation of spoiler areas and other statistics. This removed 16,378 pages from consideration, leaving us with 24,239 pages to process. Of the 24,239 pages processed, 9,167 (38%) had no mementos in the Internet Archive. Without observations, the spoiler probability is effectively 0% for these pages, but they cannot be

Table 1 Fan wikis used in the spoiler areas experiment

Television show (network)	Wiki URI	# of pages	First revision	First episode
The Big Bang Theory (CBS)	bigbangtheory.wikia.com	1120	2007-12-14	2007-09-24
Boardwalk Empire (HBO)	boardwalkempire.wikia.com	2091	2010-03-18	2010-08-23
Breaking Bad (A&E)	breakingbad.wikia.com	995	2009-04-27	2008-01-20
Continuum (Showcase)	continuum.wikia.com	258	2012-11-13	2012-05-27
Downton Abbey (BBC)	downtonabbey.wikia.com	784	2010-10-04	2010-09-26
Game of Thrones (HBO)	gameofthrones.wikia.com	3144	2010-06-24	2011-04-17
Grimm (NBC)	grimm.wikia.com	1581	2010-04-14	2011-10-28
How I Met Your Mother (CBS)	how-i-met-your-mother.wikia.com	1709	2008-07-21	2005-09-19
Lost (ABC)	lostpedia.wikia.com	18,790	2005-09-22	2004-09-22
Mad Men (AMC)	madmen.wikia.com	652	2009-07-25	2007-06-03
NCIS (CBS)	ncis.wikia.com	5345	2006-09-25	2003-09-23
Once Upon A Time (ABC)	onceuponatime.wikia.com	1470	2011-08-09	2011-10-23
Scandal (ABC)	scandal.wikia.com	331	2011-06-07	2012-04-05
True Blood (HBO)	trueblood.wikia.com	1838	2008-10-06	2008-09-07
White Collar (USA)	whitecollar.wikia.com	506	2009-10-30	2009-10-23
Total		40,617		

Table 2 Data reduction of wiki pages used in experiment

	Count
Wiki pages from all fan wikis	40,617
Pages that are just wiki redirects	16,378
Non-redirect pages with no Internet Archive observations	9,167
Wiki pages remaining for experiment	15,064

visited by our fan, so they are no use either. For these reasons these pages were removed. This left us with only 15,064 pages that actually had observations in the Internet Archive at the time the wiki exports were extracted.

Table 3 shows the statistics for each fan wiki. There is a mean overall spoiler probability for all 15,064 pages of 0.659, ranging from 0.39, for the *White Collar* wiki, to 0.818, for the *NCIS* wiki. As Steiner, et al., observed, real-world events, in this case episodes, influence fan activity, thus inspiring fans to create revisions in each wiki. The observations by the Internet Archive are driven by crawling algorithms rather than directly by real-world events or fan activity, resulting in large gaps between returns to these wiki sites, causing the number of observations to be far lower.

Figure 11 shows a cumulative distribution function of spoiler probabilities for all wikis within the data set. For more than 74% of the pages, the spoiler probability is ≥ 0.5.

The number of observations per day is an order of magnitude smaller than the number of revisions per day. Because we have a complete history of each page, wikis provide a unique opportunity to demonstrate the amount of web archive coverage for each revision. Figure 12 shows unobserved revisions. Each colored line in the graph corresponds to the lifespan of one of the 15,064 wiki pages. Darker colors indicate wiki revisions that contain no observations. The color of each line changes to white once an observation is made by the Internet Archive. The color then turns darker again as unobserved revisions are created for each wiki page. Using this visualization, we see the observation history of all pages all at once. Due to this color scheme, apparent vertical lines are generated by the shifts in color of the individual timelines. These vertical lines demonstrate trends in mass web archiving for these wiki pages, where large numbers of pages were archived at the same time in 2009, late in 2011, and late in 2013. Even though more recent pages have lighter timelines, meaning fewer unobserved revisions, the colors indicate that there are still a number of revisions without observations, demonstrating the sparseness of the holdings of the Internet Archive compared to the wikis used in our experiment.

Table 4 contains statistics for the most popular page in each of the wikis that we have surveyed, where popularity is determined by the number of page revisions generated, as inspired by the aforementioned work of Almedia, Mozafari, and Cho [3]. Seeing as these wikis are authored by fans, readers familiar with many of these television shows will not be surprised that most of the popular pages are about main characters. The table also lists the number of spoiler areas, revisions, and observations, showing how there is not a simple correlation between these values and spoiler probability.

Figure 13 shows spoiler areas for the page with the most revisions in our entire dataset, a page from *Lostpedia* about the character Kate Austen. Each spoiler area is shown in red with some degree of transparency. The probability of encoun-

Table 3 Overall statistics for each fan wiki, including mean probability of spoiler for all pages

Wiki	Probability of spoiler			Revisions/day			Observations/day		
	Mean	Std dev	Rel Err	Mean	Std dev	Rel Err	Mean	Std dev	Rel Err
bigbangtheory	0.6671	0.1603	0.0116	0.0506	0.0668	0.0639	0.0033	0.0034	0.0488
boardwalkempire	0.4173	0.1698	0.0160	0.0102	0.0185	0.0718	0.0022	0.0026	0.0452
breakingbad	0.7457	0.2053	0.0127	0.0185	0.0351	0.0872	0.0032	0.0032	0.0459
continuum	0.3935	0.1767	0.0471	0.0317	0.0250	0.0829	0.0051	0.0023	0.0479
downtonabbey	0.5848	0.1739	0.0196	0.0374	0.0636	0.1124	0.0020	0.0013	0.0419
gameofthrones	0.4734	0.2483	0.0122	0.0425	0.0652	0.0356	0.0041	0.0049	0.0279
grimm	0.4788	0.1753	0.0201	0.0700	0.0857	0.0672	0.0027	0.0015	0.0305
how-i-met-your-mother	0.7407	0.1004	0.0046	0.0163	0.0220	0.0463	0.0014	0.0010	0.0263
lostpedia	0.7678	0.1629	0.0027	0.0391	0.1083	0.0348	0.0040	0.0055	0.0173
madmen	0.5297	0.1436	0.0133	0.0049	0.0076	0.0764	0.0014	0.0021	0.0755
ncis	0.8180	0.1065	0.0041	0.0073	0.0097	0.0413	0.0009	0.0008	0.0279
onceuponatime	0.5162	0.1630	0.0132	0.1271	0.1327	0.0437	0.0037	0.0025	0.0281
scandal	0.5914	0.1648	0.0269	0.0418	0.0484	0.1120	0.0030	0.0019	0.0608
trueblood	0.5170	0.1617	0.0106	0.0210	0.0410	0.0658	0.0016	0.0016	0.0345
whitecollar	0.3902	0.2500	0.0500	0.0117	0.0147	0.0986	0.0019	0.0015	0.0609
Overall	0.6627	0.2221	0.0028	0.0360	0.0868	0.0201	0.0032	0.0044	0.0115

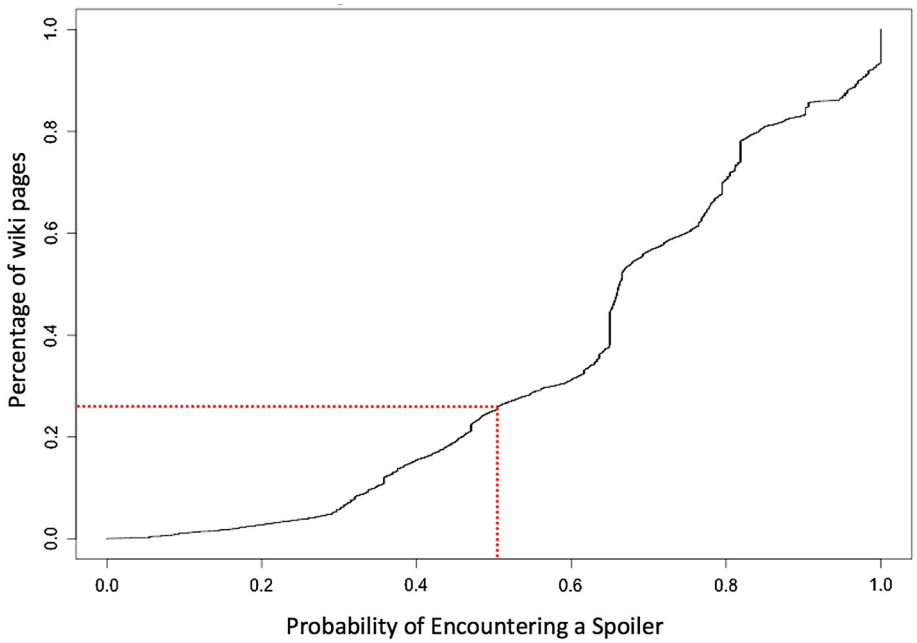

Fig. 11 *Graph* of the cumulative distribution function of spoiler probabilities for all 15 wiki sites, showing that more than 74% of pages contain a spoiler probability ≥ 0.5

tering a spoiler for Kate's page is 67%, calculated by Eq. (2). Because television shows air during different times of the year, we see collections of diamonds on the episode timeline, indicating the datetimes of seasons and partial seasons. Confirming the behavior seen by Steiner, revisions do follow episodes, and the revisions taper off after the final episode of the series. We also see the Internet Archive on the bottom line taking years to establish a regular archiving pattern, finally archiving the page on a regular basis long after the final show has aired.

Figure 14 shows spoiler areas for the page about the *Big Bang Theory* character Sheldon Cooper. Sheldon's page has a spoiler probability of only 31%. We can see the clusters of points indicating each season on the events timeline. Even though Sheldon's page contains quite a few spoiler areas after the second season, there appears to be a block of time

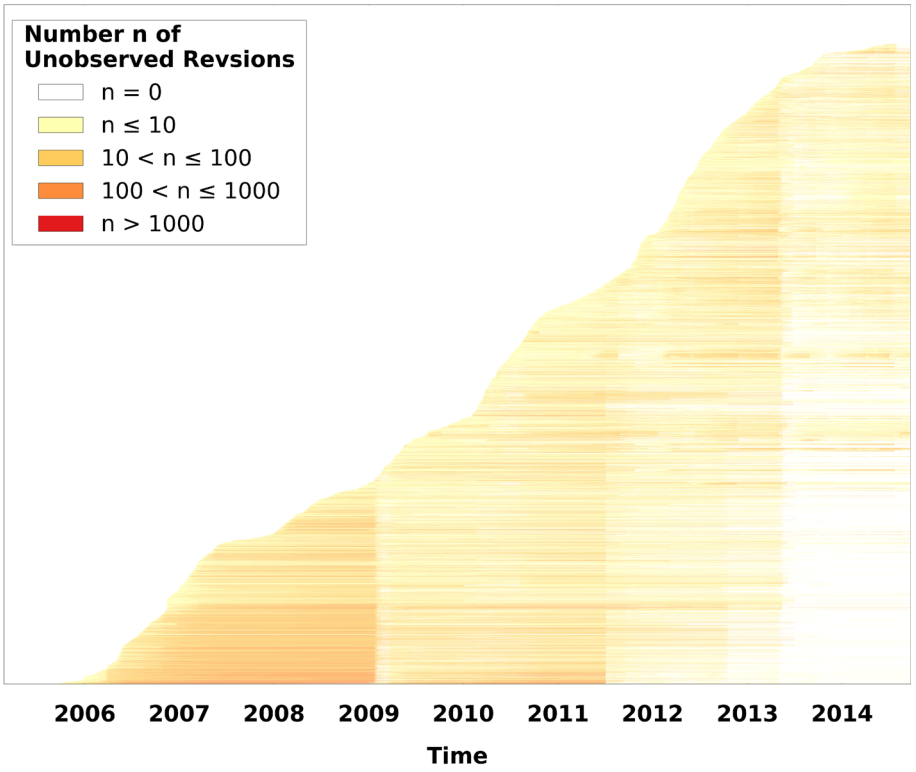

Fig. 12 Demonstration of unobserved revisions; each *colored line* is a timeline for each of the 15,064 wiki pages in our dataset; *the darker the color*, the higher the number of unobserved revisions (color figure online)

Table 4 Spoiler probabilities for the most popular pages within each fan wiki

Wiki	Page name	Probability of spoiler	# of spoiler areas	# of revisions	# of observations
bigbangtheory	Sheldon Cooper	0.31	69	1958	30
boardwalkempire	Nucky Thompson	0.15	31	290	15
breakingbad	Walter White	0.43	40	882	20
continuum	Keira Cameron	0.54	21	104	5
downtonabbey	Sybil Branson	0.42	23	580	3
gameofthrones	Daenerys Targaryen	0.16	24	768	29
grimm	Nick Burkhardt	0.39	30	795	5
how-i-met-your-mother	Barney Stinson	0.55	120	588	13
lostpedia	Kate Austen	0.67	94	3531	27
madmen	Mad Men Wiki	0.22	36	250	85
ncis	Abigail Sciuto	0.67	182	404	11
onceuponatime	Emma Swan	0.36	34	1210	11
scandal	Main Page	0.60	31	250	14
trueblood	Eric Northman	0.28	47	931	14
whitecollar	Neal Caffrey	0.29	38	199	8

before the third season where one is safe to browse this page and avoid spoilers. We see no tapering off as with Lostpedia, because this show is still airing. Unlike Lostpedia, which was started in 2005, observations for Sheldon's wiki page were started much closer to its initial revision in 2007.

Figure 15 provides another example, using a page from the *Game of Thrones* wiki, with a spoiler probability of 16%. This example really demonstrates the small annual seasons on the top timeline. We also see revisions tapering off between seasons, as there is little new content for wiki editors to contribute. From this visualization, there are many spoiler areas throughout the life of the wiki page, and spoiler-free browsing is done best between seasons.

For more visualizations, details, and additional results, please see [15].

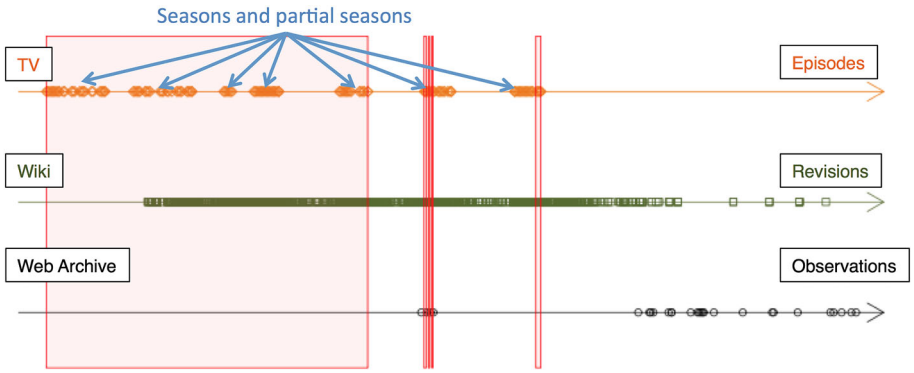

Fig. 13 Spoiler areas for the most popular page (3531 revisions) in our data set, from the *Lostpedia* wiki

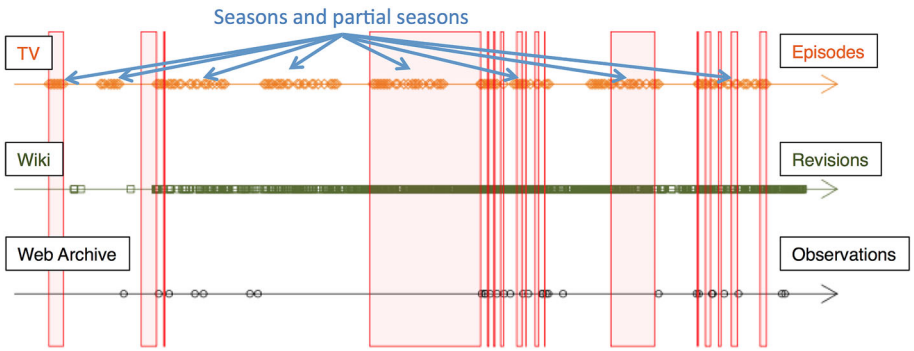

Fig. 14 Spoiler areas for the most popular page (1958 revisions) in *the Big Bang Theory* wiki

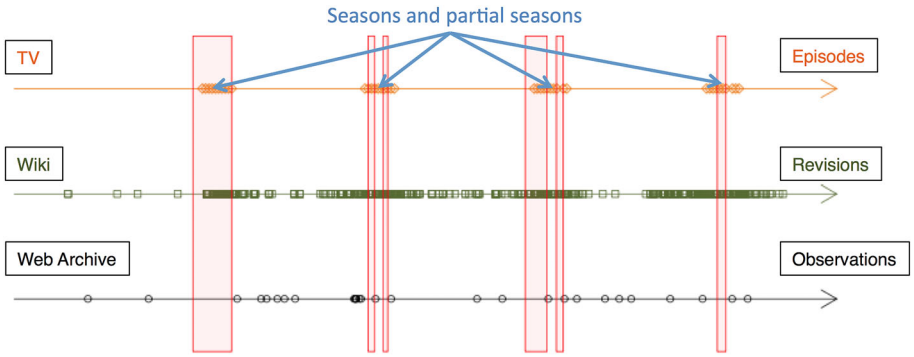

Fig. 15 Spoiler areas for the most popular page (768 revisions) in the *Game of Thrones* wiki

Using these spoiler area visualizations, our fan can determine safe datetimes to choose to avoid spoilers, but such visualizations would need to be calculated for each page as our fan visits them. Even with such work, 38% of the pages have no observations in the Internet Archive, ironically allowing the fan to avoid spoilers, but of course providing no content either.

5 Picking the best day to avoid spoilers

Our theoretical fan discovered how often one could encounter spoilers when using Memento on the Internet Archive. If this fan wanted to still use Memento on Internet Archive, which day prior to the episode is best to select as an accept-datetime value when trying to avoid spoilers? After an episode airs, fans create wiki revisions based on information from that episode. Our fan wishes to maximize the amount of information by getting the most recent content while still avoiding a spoiler. Intuition indicates that one should select an accept-datetime within 1 day prior to the episode to avoid spoilers for that episode, but how effective is this choice understanding the mindist heuristic used by the Internet Archive?

In order to determine the best choice of accept-datetime value, we conducted a second experiment. In this experiment we used Memento on the Internet Archive with the most popular pages from each of the fan wikis in our dataset. For each popular page, we then performed the following:

1. Choose an accept-datetime matching the datetime of the episode for which we are trying to avoid spoilers.
2. Use Memento with this accept-datetime and the URI of the wiki page as arguments to acquire an observation from the Internet Archive.

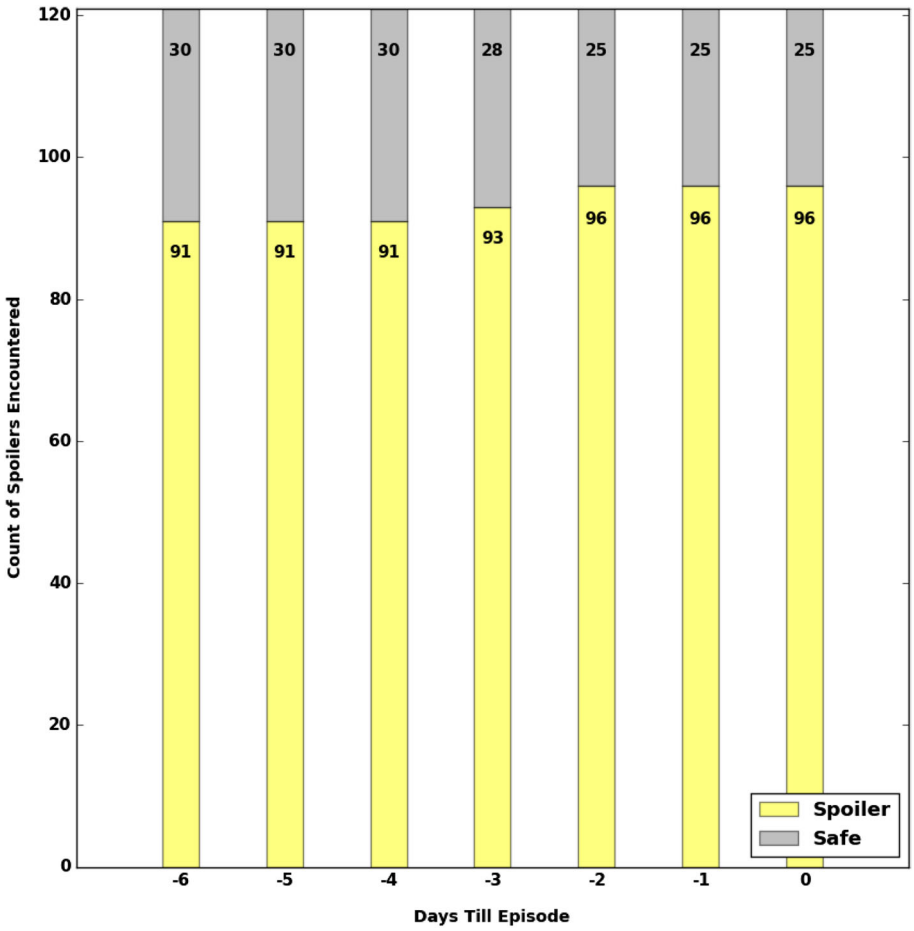

Fig. 16 *Bar chart* for most popular *Lostpedia* page, showing the ratio of spoilers to safe observations encountered when using Memento up to 6 days prior to a given episode

3. As in the first experiment, use Eq. (1) to determine if the observation is a spoiler or safe.
4. Choose an accept-datetime of 1 day prior to the datetime of the episode and repeat.
5. Choose an accept-datetime of 2 days prior and repeat again, then repeat again with 3 days prior and so on up to 6 days prior. We stop at 6 days because all series chosen are weekly series, and going back further would lead us into the week for a different episode.
6. Repeat for each episode in the series.

Figure 16 shows the results in a stacked bar chart for the most popular page in our dataset, from *Lostpedia*. Yellow is used to indicate how many spoilers were encountered for the accept-datetime chosen. Gray is the number of safe observations. For this page we see that between 4 and 6 days prior to an episode is safest for avoiding spoilers. We also see that the majority of observations did result in spoilers nonetheless.

In Fig. 17 we see that choosing 6 days prior to an episode, meaning 1 day after the prior episode, is safest for the most popular *Big Bang Theory* wiki page. Interestingly enough, for Sheldon's page, we see that the majority of observations did not result in spoilers.

The most popular *Game of Thrones* wiki page from Fig. 18 demonstrates that between 5 and 6 days prior to an episode is safest for avoiding spoilers. Unlike the other two pages, these first 2 days have an almost even chance of encountering spoilers, but of course, that chance goes up as one gets closer to the episode date.

All of the results for accept-datetime choice are aggregated from the 15 popular pages and visualized in Fig. 19 which illustrates that choosing 1 day after the prior episode is safest. Unfortunately, there are still spoilers in many cases, no matter which day is chosen.

6 Avoiding spoilers in fan wikis

A wiki records the version history of its pages as a list of revisions. But, as mentioned in the Introduction, using that history as a means to avoid spoilers is:

– *Cumbersome* It entails scrolling through the history in search of a date prior to the first airing of an episode, and then selecting a revision.
– *Spoiler-prone* Although the selected revision will not contain spoilers, wiki pages linked from it will, because

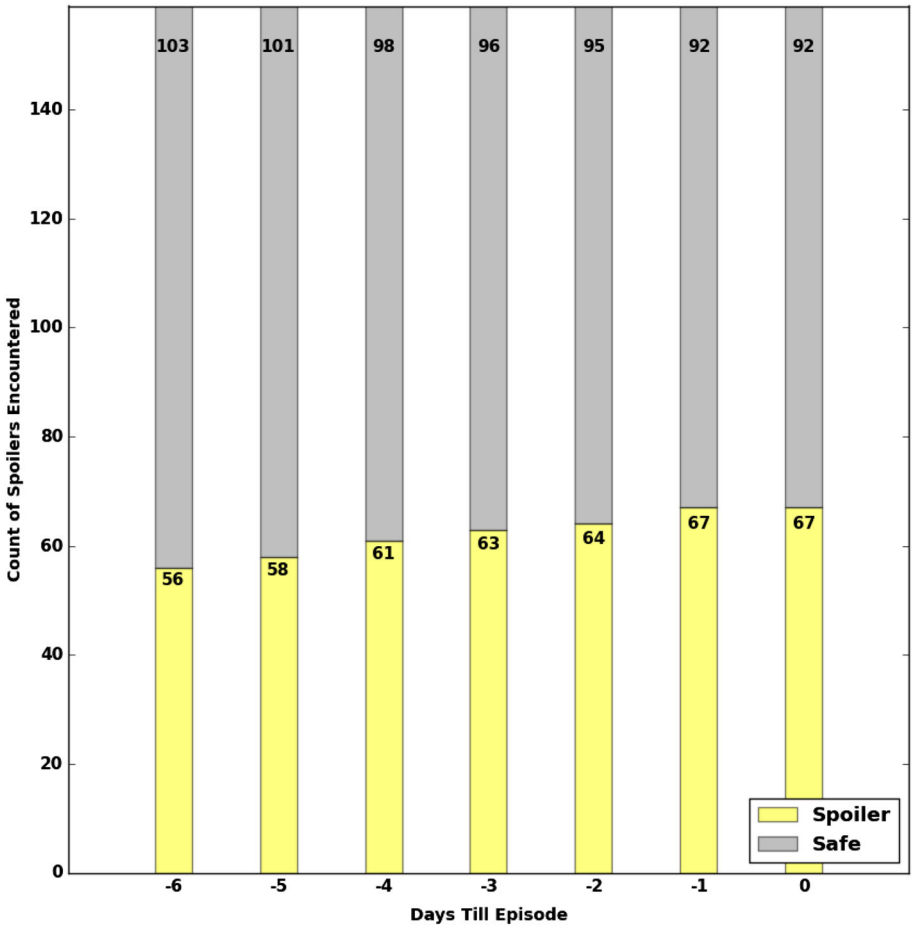

Fig. 17 *Bar chart* for most popular *Big Bang Theory* wiki page, showing the ratio of spoilers to safe observations encountered when using Memento up to 6 days prior to a given episode

links lead to the current revision. As such, cross-page navigation has to be avoided and safely visiting multiple pages requires scrolling through the history of every desired page.

Wikis are the densest archives possible for their own resources. Hence, in order to determine which revision was operational at a given datetime, it suffices to select the revision that was created most recently prior to a desired datetime. Wikis must use minpast. As such, they are able to exactly provide the revision of a wiki page that was operational for a given accept-datetime by using the minpast (rather than mindist) heuristic.

We have developed two extensions for MediaWiki wikis that support the Memento protocol using the minpast heuristic [16]. Combined with a Memento-compliant client such as Memento for Chrome,[13] which allows selecting a preferred datetime for navigating the past of the web, this offers a robust solution to the fan wiki spoiler problem. As is the case with the web archive approach, a fan still needs to know the date of first airing of an episode. But, understanding the mechanics of the minpast heuristic, selection of any datetime prior to that airing date is spoiler safe. The closer the selected datetime is to that of airing, the more informative the resulting revision will be with regard to the previous episode. Once a user selects a datetime, a Memento client will issue each new request with the same datetime in its accept-datetime header. Because of this behavior, Memento clients also allow fans to stay in the past, continuing to browse linked web pages with this same datetime, thus avoiding spoilers on other pages because the Memento MediaWiki extension uses the minpast heuristic.

We have created a video that demonstrates the use of Memento for Chrome to access web archives and Wikipedia.[14] We have also created another video demonstrating how Memento for Chrome and the Memento MediaWiki extension can be used together to avoid spoilers.[15]

[13] https://chrome.google.com/webstore/detail/memento-time-travel/jgbfpjledahoajcppakbgilmojkaghgm?hl=en.

[14] https://www.youtube.com/watch?v=WtZHKeFwjzk.

[15] https://www.youtube.com/watch?v=ciClYjTnscs.

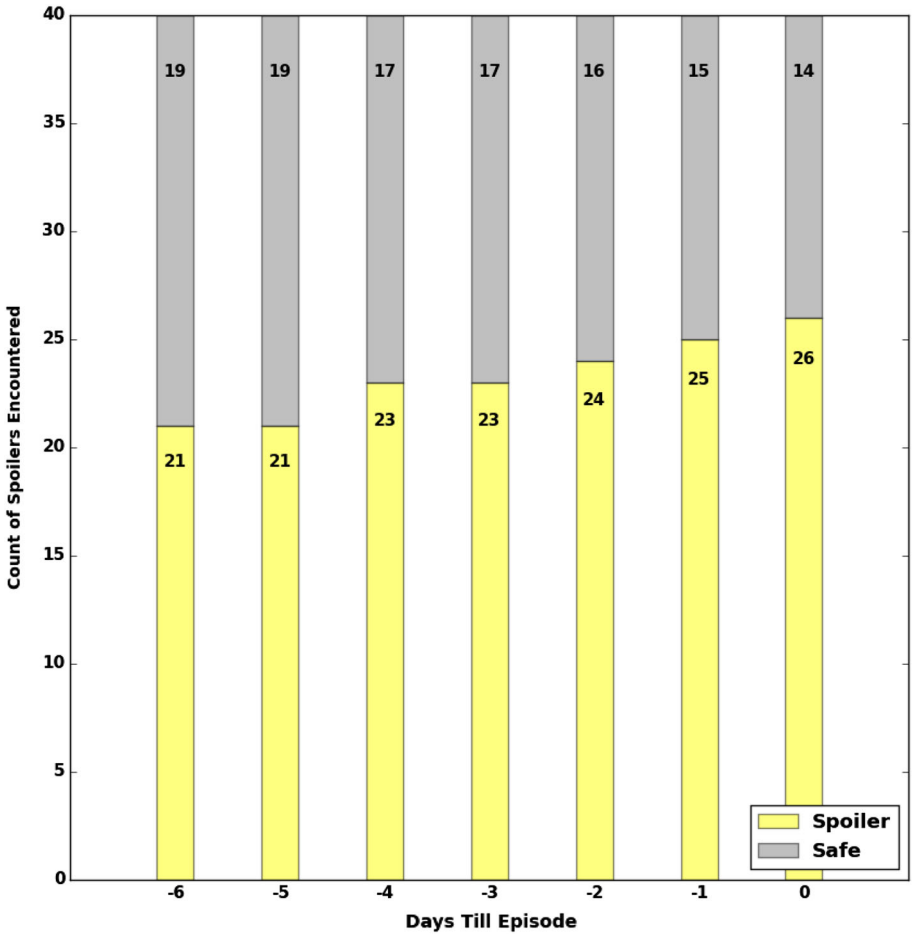

Fig. 18 *Bar chart* for most popular *Game of Thrones* wiki page showing the ratio of spoilers to safe observations encountered when using Memento up to 6 days prior to a given episode

7 Future work

It is possible to use minpast not just for avoiding spoilers in television shows, but potentially sporting events. Sports fans who have not seen, but have recorded, a game can use minpast to avoid discovering scores and highlights until they have caught up. A study can be conducted as to how well minpast will work for this use case, potentially using versioned resources for baseball[16] or the Olympics.[17]

Researchers into historical context may seek to avoid information on emerging topics. Minpast can be used to discover historical discussions on topics with evolving information such as same-sex marriage[18] or presidential elections.[19] They can answer questions such as "knowing what we know now, what did people understand at the given time?" or "not knowing what we know now, what did people think at the given time?" Consider the use case of documenting attitudes towards terrorism in the United States on September 10, 2011, 1 day prior to terrorist attacks that killed almost 3,000 people. It is also possible to use minpast programmatically to extract information from wiki edits, determining not only when a specific event occurred, but what change happened.

For resources without access to all page revisions, such as sparse web archives, one can compare past and future mementos to determine if spoilers have been revealed. One could also combine our structural approach, that is purely based on time, with one or more of the content-based natural language processing or machine learning approaches from our Related Work Section. The structural approach would provide a selection of revisions whose content could then be processed to determine their spoiler status.

8 Conclusions

Clearly, our fan who is trying to avoid spoilers by visiting the Internet Archive will be met with frustration.

We have explored how effective it is to use web archives as a means to avoid spoilers in fan wikis. We have pointed

[16] http://baseball.wikia.com/wiki/Main_Page.

[17] http://olympics.wikia.com/wiki/Olympics_Wiki.

[18] https://en.wikipedia.org/wiki/Same-sex_marriage_law_in_the_United_States_by_state.

[19] https://en.wikipedia.org/wiki/United_States_presidential_election,_2016.

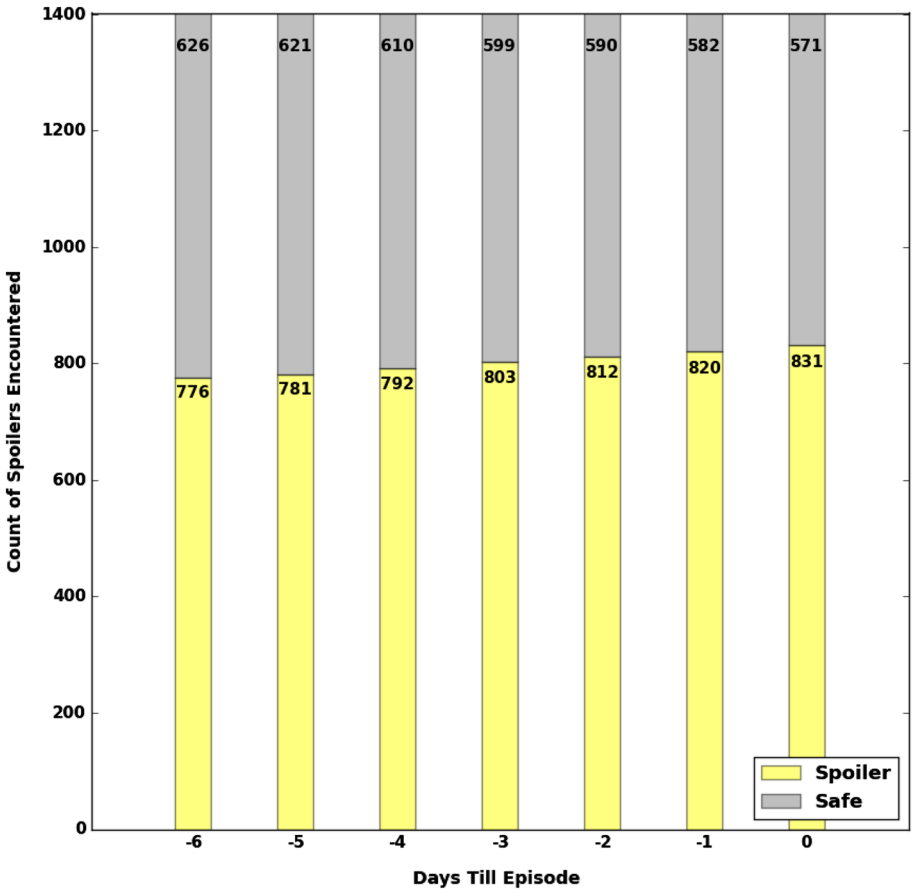

Fig. 19 *Bar chart* for all popular wiki pages showing the ratio of spoilers to safe observations encountered when using Memento up to 6 days prior to a given episode

out that web archives record observations and use a mindist heuristic. As such, for a given preferred datetime in the past, a web archive delivers the observation that is temporally closest to that datetime irrespective of whether it is before or after the preference. We found that for roughly 38% of the pages that we considered no observations were available in the Internet Archive at all, and more than 74% of the pages have a 50% or higher chance of including a spoiler. We also investigated what the safest day in the week prior to an episode is to select as datetime to search a web archive. Counter to our personal intuition, likely guided by a desire to have maximum information about the prior episode, in most cases it is not the day prior to the episode but rather the day after the previous episode.

We have also pointed out that fan wikis have all the information required to address the spoiler problem autonomously, because they record full page histories instead of distinct observations of pages. These wikis are indeed dense web archives in their own right. By leveraging their own history and supporting the Memento protocol using a minpast heuristic, wikis can indeed provide the version of a wiki page operational at a given time and address the spoiler problem locally. This solution would allow fans equipped with a Memento client to safely navigate the wiki as long as they set their time travel date prior to that of the first airing of an episode they have not yet viewed. We speculate that this capability would make quite a few fans happy.

Acknowledgements This work was made possible in part by a Grant from the Andrew Mellon Foundation. We are grateful to the Internet Archive for their continued support of Memento. Memento is a joint project between the Los Alamos National Laboratory Research Library and Old Dominion University.

References

1. Ainsworth, S.G., Nelson, M.L.: Evaluating sliding and sticky target policies by measuring temporal drift in acyclic walks through a web archive. JCDL, 39–48 (2013). doi:10.1145/2467696.2467718
2. Ainsworth, S.G., Nelson, M.L., Van de Sompel, H.: A Framework for Evaluation of Composite Memento Temporal Coherence. Tech. Rep., Old Dominion University (2014) arXiv:1402.0928
3. Almeida, R., Mozafari, B., Cho, J.: On the Evolution of Wikipedia. In: ICWSM (2007). http://www.icwsm.org/papers/paper2.html. Accessed 03 May 2014
4. AlNoamany, Y., Weigle, M.C., Nelson, M.L.: Detecting off-topic pages within timemaps in web archives. Int J Digit. Libr., 1–19 (2016). doi:10.1007/s00799-016-0183-5
5. AlNoamany, Y.A., Weigle, M.C., Nelson, M.L.: Access patterns for robots and humans in web archives. In: JCDL, 339–348 (2013). doi:10.1145/2467696.2467722

6. Boyd-Graber, J., Glasgow, K., Zajac, J.S.: Spoiler alert: machine learning approaches to detect social media posts with revelatory information. ASIS&T **50**, 1–9 (2013). doi:10.1002/meet.14505001073
7. Cohen, N.: Spoiler alert: whodunit? Wikipedia will tell you. New York Times (2010). http://www.nytimes.com/2010/09/18/business/media/18spoiler.html. Accessed 16 Sep 2014
8. Cuslidge, T., Weiss, J.: Potter fans' new foe? The Web. Pop Matters (2007). http://www.popmatters.com/article/potter-fans-new-foe-the-web/. Accessed 04 Oct 2014
9. Denham, J.: Netflix releases house of cards 'Spoiler Foiler' for twitter users. The Independent (2014). http://www.independent.co.uk/arts-entertainment/tv/news/netflix-releases-house-of-cards-spoiler-foiler-for-twitter-users-9136324.html. Accessed 24 Oct 2014
10. Golbeck, J.: The twitter mute button: a web filtering challenge. In: SIGCHI, 2755–2758 (2012). doi:10.1145/2207676.2208673
11. Gross, B.D.: Spoiler alert! Negotiating social media in the DVR age. CNN (2014). http://www.cnn.com/2014/02/25/tech/social-media/spoilers-social-media/. Accessed 24 Oct 2014
12. Hart, H.: Spoiler wars heat up as lost returns. Wired (2009). https://www.wired.com/2009/01/new-lost-season/. Accessed 16 Sep 2014
13. Jeon, S., Kim, S., Yu, H.: Don't be spoiled by your friends: spoiler detection in TV program tweets. In: Seventh International AAAI Conference on Weblogs and Social Media, pp. 681–684 (2013). http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/5979. Accessed 02 Nov 2015
14. Johns, M.D.: Two screen viewing and social relationships: exploring the invisible backchannel of TV viewing. In: CATTaC, pp. 333–343 (2012). http://sammelpunkt.philo.at:8080/2159/. Accessed 16 Sep 2014
15. Jones, S.M.: Avoiding Spoilers on MediaWiki Fan Sites Using Memento. Master's thesis, Old Dominion University (2015). http://digitalcommons.odu.edu/computerscience_etds/1/
16. Jones, S.M., Nelson, M.L., Shankar, H., Van de Sompel, H.: Bringing Web Time Travel to MediaWiki: An Assessment of the Memento MediaWiki Extension. Tech. Rep., Old Dominion University (2014). arXiv:1406.3876
17. Leaver, T.: Watching Battlestar Galactica in Australia and the Tyranny of Digital Distance. Media International Australia, Incorporating Culture and Policy (126), 145 (2008). doi:10.1177/1329878X0812600115. http://mia.sagepub.com/content/126/1/145.abstract
18. Leavitt, J.D., Christenfeld, N.J.S.: Story spoilers don't spoil stories. Psychol Sci **22**(9), 1152–1154 (2011). doi:10.1177/0956797611417007
19. Leuf, B., Cunningham, W.: The Wiki Way: Quick Collaboration on the Web. Addison-Wesley Longman Publishing Co. Inc, Boston (2001)
20. Mittell, J.: Sites of participation: Wiki fandom and the case of Lostpedia. TWC (2009). doi:10.3983/twc.2009.0118
21. SalahEldeen, H.M., Nelson, M.L.: Losing My Revolution: How Many Resources Shared on Social Media Have Been Lost?, pp. 125–137. Springer, Berlin (2012). doi:10.1007/978-3-642-33290-6_14
22. Schirra, S., Sun, H., Bentley, F.: Together alone. SIGCHI, pp. 2441–2450 (2014). doi:10.1145/2556288.2557070
23. Steel, E.: Those Dreaded Spoilers That Can Torpedo Dramatic Plot Take On a New Meaning. New York Times (2014). http://www.nytimes.com/2014/09/22/business/media/those-dreaded-spoilers-that-can-torpedo-dramatic-plot-take-on-a-new-meaning.html. Accessed 16 Sep 2014
24. Steiner, T., van Hooland, S., Summers, E.: MJ No More: Using Concurrent Wikipedia Edit Spikes with Social Network Plausibility Checks for Breaking News Detection. In: Proceedings of the 22nd International Conference on World Wide Web, ACM, pp. 791–794 (2013). doi:10.1145/2487788.2488049
25. Toyoda, M., Kitsuregawa, M.: The History of Web Archiving. In: Proceedings of the IEEE, vol. 100 (Special Centennial Issue), pp. 1441–1443 (2012). doi:10.1109/JPROC.2012.2189920
26. Tsang, A.S.L., Yan, D.: Reducing the Spoiler Effect in Experiential Consumption. Advances in Consumer Research (36), 708–709 (2009). http://www.acrwebsite.org/volumes/14341/volumes/v36/NA-36. Accessed 10 Feb 2015
27. Van de Sompel, H., Nelson, M.L., Sanderson, R.: RFC 7089: HTTP Framework for Time-Based Access to Resource States—Memento (2013). http://tools.ietf.org/rfc/rfc7089.txt. Accessed 03 May 2014
28. Van de Sompel, H., Nelson, M.L., Sanderson, R., Balakireva, L., Ainsworth, S., Shankar, H.: Memento: Time travel for the web. Tech. Rep., Los Alamos National Laboratories and Old Dominion University (2009). arXiv:0911.1112
29. Van de Sompel, H., Sanderson, R., Nelson, M., Balakireva, L., Shankar, H., Ainsworth, S.: An HTTP-based versioning mechanism for linked data. In: LDOW (2010). http://events.linkeddata.org/ldow2010/papers/ldow2010_paper13. Accessed 08 Aug 2016
30. Yaspan, A.: Essentials of Probability. Prindle, Weber and Schmidt, Boston (1968)
31. Zittrain, J., Albert, K., Lessig, L.: Perma: scoping and addressing the problem of link and reference rot in legal citations. Leg. Inf. Manag. **14**(2), 88–99 (2014). doi:10.1017/S1472669614000255