Avoiding Spoilers On MediaWiki Fan Sites Using Memento

By

Shawn M. Jones

sjone@cs.odu.edu

Committee:
Dr. Michael L. Nelson
Dr. Michele C. Weigle
Dr. Irwin B. Levinstein
Warning: This presentation may contain spoilers
It Started With A Discussion At Work About This Guy From *Game Of Thrones*
So We Use Fan Wikis, Because They Are Useful For Our Discussions

http://gameofthrones.wikia.com/wiki/Joffrey_Baratheon
Joffrey Baratheon is a major character in the first, second, third, and fourth seasons. He is played by Jack Gleeson and debuts in the series premiere. Joffrey ruled the Seven Kingdoms, having claimed the Iron Throne after his legal father, King Robert Baratheon, was killed. Joffrey was the bastard son of the incestuous relationship between his father and his Queen Cersei Lannister, who was forcibly raped by the Iron Throne war lord and future king, Stannis Baratheon. Joffrey was the heir to the Houses Baratheon and Lannister, the First of His Name, King of the Andals and the First Men, Lord of the Seven Kingdoms, and Protector of the Realm.

Joffrey was originally betrothed to Sansa Stark, whom he kept captive in King's Landing after he beheaded her father for treason, as Eddard Stark tried to reveal the truth about his parentage. However, he later set Sansa aside in favor of Margaery Tyrell, as part of the new alliance between House Lannister, the primary supporters of Joffrey, and House Tyrell, the ruling house of the Reach whose support helped save King's Landing from Stannis Baratheon's attack.

At his wedding to Margaery Tyrell, Joffrey was assassinated by drinking poisoned wine. His final gesture as he was dying was towards his uncle, Tyrion Lannister. Upon his death, Cersei Lannister arrests Tyrion on the grounds of regicide, though the actual orchestrators are Lord Petyr Baelish and Lady Olenna Tyrell.
Joffrey Baratheon is a major character in the first, second, third, and fourth seasons. He is played by Jack Gleeson and debuts in the series premiere. Joffrey ruled the Seven Kingdoms, having claimed the Iron Throne after his legal father, King Robert Baratheon, was killed in a battle. As the oldest son of the incestuous King Robert and Queen Cersei, Joffrey was born with the iron stains on his face that give him his name. He was the first of his name, King of the Andals and the First Men, Lord of the Seven Kingdoms, and Protector of the Realm.

Joffrey was originally betrothed to Sansa Stark, whom he kept captive in King’s Landing after he beheaded her father for treason, as Eddard Stark tried to reveal the truth about his parentage. However, he later set Sansa aside in favor of Margaery Tyrell, as part of the new alliance between House Lannister, the primary supporters of Joffrey, and House Tyrell, the ruling house of the Reach whose support helped save King’s Landing from Stannis Baratheon’s attack.

At his wedding to Margaery Tyrell, Joffrey was assassinated by drinking poisoned wine. His final gesture as he was dying was towards his uncle, Tyrion Lannister. Upon his death, Cersei Lannister arrests Tyrion on the grounds of regicide, though the actual orchestrators are Lord Petyr Baelish and Lady Olenna Tyrell.
Joffrey Baratheon is a major character in the first, second, third, and fourth seasons. He is played by Jack Gleeson and debuts in the series premiere. Joffrey ruled the Seven Kingdoms, having claimed the Iron Throne after his legal father, King Robert Baratheon, died. He was the bastard son of the incestuous relationship between Robert Baratheon and his sister, Queen Cersei Lannister. Joffrey was the last ruler of the House Baratheon and the first of his name, King of the Andals and the First Men, Lord of the Seven Kingdoms, and Protector of the Realm.

Joffrey was originally betrothed to Sansa Stark, whom he kept captive in King's Landing after he beheaded her father for treason, as Eddard Stark tried to reveal the truth about his parentage. However, he later set Sansa aside in favor of Margaery Tyrell, as part of the new alliance between House Lannister, the primary supporters of Joffrey, and House Tyrell whose support helped save Baratheon's attack.

At his wedding to Margaery Tyrell, he succumbed to drinking poisoned wine. His final gesture as he was dying was towards his uncle, Tyron Lannister. Upon his death, Cersei Lannister seizes Tyrion on the grounds of regicide, though the actual orchestrators are Lord Petyr Baelish and Lady Olenna Tyrell.

He was assassinated

<table>
<thead>
<tr>
<th>Status</th>
<th>Deceased</th>
</tr>
</thead>
</table>

Motivation
But The Current Page For This Character Contains Spoilers

Joffrey Baratheon is a major character in the first, second, third, and fourth seasons. He is played by Jack Gleeson and debuts in the series premiere. Joffrey ruled the Seven Kingdoms, having claimed the Iron Throne after his legal father, King Robert Baratheon, was killed by a plot against him. Joffrey was the Cersei Lannister and Tywin Lannister's bastard son and the father of Myrcella Baratheon.

He was the eldest son of the Houses Baratheon and Lannister, the First of His Name, King of the Andals and the First Men, Lord of the Seven Kingdoms, and Protector of the Realm.

Joffrey was originally betrothed to Sansa Stark, whom he kept captive in King's Landing after he beheaded her father for treason, as Eddard Stark tried to reveal the truth about his parentage. However, he later set Sansa aside in favor of Margaery Tyrell, as part of the new alliance between House Lannister, the primary supporters of Joffrey, and House Tyrell, whose support helped save Baratheon's attack.

At his wedding to Margaery Tyrell, Joffrey was assassinated with a poisoned glass of wine. Upon his death, Cersei Lannister was accused of poisoning him with the grounds of regicide, though the Tyrells held him responsible. Petyr Baelish and Lady Olenna were implicated as well.

Deceased
But The Current Page For This Character Contains Spoilers

Joffrey Baratheon is a major character in the first, second, third, and fourth seasons. He is played by Jack Gleeson and debuts in the series premiere. Joffrey ruled the Seven Kingdoms, having claimed the Iron Throne after his legal father, King Robert Baratheon, was assassinated. He was the bastard son of the incestuous relationship between his father and Queen Cersei Lannister. Joffrey was born in King's Landing and is the first in line to the Iron Throne.

Joffrey was originally betrothed to Sansa Stark, whom he kept captive in King's Landing. He beheaded her father for treason, as Eddard Stark tried to reveal the truth about his parentage. However, he later set Sansa aside in favor of Margery Tyrell, as part of the new alliance between House Lannister, the primary supporters of Joffrey, and House Tyrell of the Reach whose support helped save the Baratheon's attack.

At his wedding to Margery Tyrell, Joffrey was poisoned during a toast to his new bride. The primary cause of death was his consumption of poisoned wine. Upon his death, the Baratheon line died out, and the Iron Throne passed to Stannis Baratheon, a cousin of Joffrey's.
We All Enjoy Some Episodic Fiction So Much...
...That Fans Have Created Wikis...
...And The Rest Of Us Read Them
So, What If We Could Avoid The Spoilers By Using Past Wiki Pages?

"We’ve had vicious kings, and we’ve had idiot kings... but I don’t know if we’ve ever been cursed with a vicious idiot for a king!"
—Tyrion Lannister, summarizing King Joffrey

Joffrey Baratheon is a major character in the first, second and third seasons. He is played by Jack Gleeson and debuts in the series premiere. Joffrey is the current ruler of the Seven Kingdoms, having claimed the Iron Throne after his father, King Robert Baratheon, died. However Joffrey is actually the bastard son of the incestuous relationship between Ser Jaime Lannister and Queen Cersei Lannister. Because of this, his claim to the Iron Throne is challenged by Robert’s younger brother, Stannis Baratheon. He is formally styled as His Grace, Joffrey of the Houses Baratheon and Lannister, the First of His Name, King of the Andals and the First Men, Lord of the Seven Kingdoms, and Protector of the Realm.

Joffrey was originally betrothed to Sansa Stark, whom he kept captive in King’s Landing after he beheaded her father for treason, as Eddard Stark tried to reveal the truth about his parentage. However, he later set Sansa aside in favor of Margaery Tyrell, as part of the new alliance between House Lannister, the primary supporters of Joffrey, and House Tyrell, the ruling house of the Reach whose support helped save King’s Landing from Stannis Baratheon.

Order Of Discussion

- Background
- Related Work
- TimeGate Heuristics
- Theory Of Spoiler Probability
- Measurements Of Spoiler Probability
- Spoilers In The Wayback Machine
- The Memento MediaWiki Extension
- Future Work
- Conclusions
Building To The Naïve Spoiler Concept

BACKGROUND

• Background
• Related Work
• TimeGate Heuristics
• Theory of Spoiler Probability
• Measurements of Spoiler Probability
• Spoilers In The Wayback Machine
• The Memento MediaWiki Extension
• Future Work
• Conclusions
Most Of Us Are Familiar With The Web Browser And HTML

The Web Browser is how we view web pages

HTML is what the browser parses to render the page
Hypertext Transfer Protocol (HTTP) is what delivers pages to the browser.
Here Is An Example HTTP Request From Google Chrome

```
1 GET /wiki/The_Hunger_Games HTTP/1.1
2 Host: en.wikipedia.org
3 Accept: image/webp, */*; q=0.8
4 Accept-Encoding: gzip, deflate, sdch
5 Accept-Language: en-US, en; q=0.8
6 Referer: https://plus.google.com/
7 User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/36.0.1985.143 Safari/537.36
```
And The Interesting Parts Of Our Request...

Get me /wiki/The_Hunger_Games from en.wikipedia.org
Here is an example HTTP Response from Wikipedia’s server for that page:

```
HTTP/1.1 200 OK
Accept-Ranges: bytes
Age: 352057
Cache-Control: private, s-maxage=0, max-age=0, must-revalidate
Connection: keep-alive
Content-Encoding: gzip
Content-language: en
Content-Length: 13769
Content-Type: text/html; charset=UTF-8
Date: Fri, 15 Aug 2014 13:43:03 GMT
Server: Apache
Set-Cookie: GeoIP=US; Norfolk:36.9312:-76.2397:v4; Path=/; Domain=.wikipedia.org
Vary: Accept-Encoding, Cookie
Via: 1.1 varnish, 1.1 varnish
X-Cache: cp1052 hit (4), cp1065 frontend hit (906)
X-Content-Type-Options: nosniff
X-UA-Compatible: IE=Edge
X-Varnish: 350285837 350035840, 1753471343 1233627904

... entity begins here
```
And The Interesting Parts Of Our Response...

<table>
<thead>
<tr>
<th>Line</th>
<th>HTTP/1.1 200 OK</th>
<th>OK, I Have What You Are Looking For</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Accept-Ranges: bytes</td>
<td>OK, I Have What You Are Looking For</td>
</tr>
<tr>
<td>3</td>
<td>Age: 352057</td>
<td>OK, I Have What You Are Looking For</td>
</tr>
<tr>
<td>4</td>
<td>Cache-Control: private, s-maxage=0, max-age=0, must-revalidate</td>
<td>OK, I Have What You Are Looking For</td>
</tr>
<tr>
<td>5</td>
<td>Connection: keep-alive</td>
<td>OK, I Have What You Are Looking For</td>
</tr>
<tr>
<td>6</td>
<td>Content-Encoding: gzip</td>
<td>OK, I Have What You Are Looking For</td>
</tr>
<tr>
<td>7</td>
<td>Content-language: en</td>
<td>OK, I Have What You Are Looking For</td>
</tr>
<tr>
<td>8</td>
<td>Content-Length: 13769</td>
<td>This Is How Big It Is</td>
</tr>
<tr>
<td>9</td>
<td>Content-Type: text/html; charset=UTF-8</td>
<td>This Is When It Was Last Changed</td>
</tr>
<tr>
<td>10</td>
<td>Date: Fri, 15 Aug 2014 13:43:03 GMT</td>
<td>Newline Indicating Start Of HTML</td>
</tr>
<tr>
<td>12</td>
<td>Server: Apache</td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
<tr>
<td>13</td>
<td>Set-Cookie: GeoIP=US:Norfolk:36.9312:-76.2397:v4; Path=/; Domain=.wikipedia.org</td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
<tr>
<td>14</td>
<td>Vary: Accept-Encoding, Cookie</td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
<tr>
<td>15</td>
<td>Via: 1.1 varnish, 1.1 varnish</td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
<tr>
<td>16</td>
<td>X-Cache: cp1052 hit (4), cp1065 frontend hit (906)</td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
<tr>
<td>17</td>
<td>X-Content-Type-Options: nosniff</td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
<tr>
<td>18</td>
<td>X-UA-Compatible: IE=Edge</td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
<tr>
<td>19</td>
<td>X-Varnish: 350285837 350035840, 1753471343 1233627904</td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
<tr>
<td>20</td>
<td>... entity begins here</td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>HTML Starts Here And You Know How Big It Is, So Stop Reading When You Get 13769 Bytes</td>
</tr>
</tbody>
</table>
Introducing Web Architecture: Resources Can Have Many Representations

URI: http://www.example.com/buffy/

Resource:
Article about the character Buffy the Vampire Slayer

English XHTML Representation:
metadata:
  Content-Language: en
  Content-Type: application/xhtml+xml

```
<!DOCTYPE html PUBLIC ""..."
<html xmlns="http://www....">
<head>
<title>Character: Buffy</title>
...</head>
</html>
```

Spanish HTML Representation:
metadata:
  Content-Language: es
  Content-Type: text/html

```
<!DOCTYPE html PUBLIC ""..."
<html>
<head>
<title>Carácter: Buffy</title>
...</head>
</html>
```

Chinese PDF Representation:
metadata:
  Content-Language: zh
  Content-Type: application/pdf

```
%PDF-1.4
%<Content>5 0 obj
<</Length 6 0 R/Filter /FlateDecode>>
stream
...</Content>
```
In 1996 Tim Berners-Lee discussed different dimensions of representation.

<table>
<thead>
<tr>
<th>Time</th>
<th>A resource may vary with time. For example, &quot;The Wall Street Journal&quot; varies with time. Each issue is a time-specific resource, which does not change with time. Most home pages on the Web change with time, in a less periodic way.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>When a document is translated, it is useful to be able to refer to it either in the generic, or to a particular specific translation.</td>
</tr>
<tr>
<td>Content-Type</td>
<td>A given resource may have many ways in which it can be represented on the wire, using different content-types (in HTTP terms). As an example, an image may be represented in PNG or GIF format.</td>
</tr>
<tr>
<td>Target medium</td>
<td>A given resource may be targeted specifically to a specific medium, such as a printer, being displayed on laptop screen, being displayed on a cellphone, or being projected onto a large screen for an audience. (This is currently available for selecting CSS stylesheets, but is not done at the HTTP content negotiation level)</td>
</tr>
</tbody>
</table>

[http://www.w3.org/DesignIssues/Generic.html](http://www.w3.org/DesignIssues/Generic.html)
These Representations Are Requested And Identified By HTTP Headers

<table>
<thead>
<tr>
<th>Request Header</th>
<th>Response Header</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept</td>
<td>Content-Type</td>
<td>Content type of Representation</td>
</tr>
<tr>
<td>Accept-Language</td>
<td>Content-Language</td>
<td>Language of Representation</td>
</tr>
<tr>
<td>Accept-Encoding</td>
<td>Content-Encoding</td>
<td>Medium of Representation</td>
</tr>
<tr>
<td>Accept-Charset</td>
<td>Content-Type</td>
<td>Character Set of Representation</td>
</tr>
</tbody>
</table>

These Headers Were Specified In RFC 2616

Their Updated Specification Is In RFC 7231
These Representations Are Requested And Identified By HTTP Headers

<table>
<thead>
<tr>
<th>Request Header</th>
<th>Response Header</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept</td>
<td>Content-Type</td>
<td>Content type of Representation</td>
</tr>
<tr>
<td>Accept-Language</td>
<td>Content-Language</td>
<td>Language of Representation</td>
</tr>
<tr>
<td>Accept-Encoding</td>
<td>Content-Encoding</td>
<td>Medium of Representation</td>
</tr>
<tr>
<td>Accept-Charset</td>
<td>Content-Type</td>
<td>Character Set of Representation</td>
</tr>
</tbody>
</table>

The term for this is **Content Negotiation**

It works so well that many are unaware of its ubiquity

Content Negotiation Does Not Create Representations, It Only Directs The User To Ones That Already Exist...
Using Our Headers Example, Chrome Tells The Server What You Want

```
GET /wiki/The_Hunger_Games HTTP/1.1
Host: en.wikipedia.org
Accept: image/webp, */*; q=0.8
Accept-Encoding: gzip, deflate, sdch
Accept-Language: en-US, en; q=0.8
Referer: https://plus.google.com/
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/36.0.1985.143 Safari/537.36
```

I prefer image/webp, and I want it compressed with gzip, and I want it in US English.
Using Our Headers Example, Wikipedia Tells Chrome What It Returns

```
HTTP/1.1 200 OK
Accept-Ranges: bytes
Age: 352057
Cache-Control: private, s-maxage=0, max-age=0, must-revalidate
Connection: keep-alive
Content-Encoding: gzip
Content-language: en
Content-Length: 13769
Content-Type: text/html; charset=UTF-8
Date: Fri, 15 Aug 2014 13:43:03 GMT
Server: Apache
Set-Cookie: GeoIP=US;Norfolk:36.9312:-76.2397:v4; Path=/; Domain=.wikipedia.org
Vary: Accept-Encoding, Cookie
Via: 1.1 varnish, 1.1 varnish
X-Cache: cp1052 hit (4), cp1065 frontend hit (906)
X-Content-Type-Options: nosniff
X-UA-Compatible: IE=Edge
X-Varnish: 350285837 350035840, 1753471343 1233627904

... entity begins here
```
Memento Finally Completes The Set By Including Time

<table>
<thead>
<tr>
<th>Request Header</th>
<th>Response Header</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept</td>
<td>Content-Type</td>
<td>Content type of Representation</td>
</tr>
<tr>
<td>Accept-Language</td>
<td>Content-Language</td>
<td>Language of Representation</td>
</tr>
<tr>
<td>Accept-Encoding</td>
<td>Content-Encoding</td>
<td>Medium of Representation</td>
</tr>
<tr>
<td>Accept-Charset</td>
<td>Content-Type</td>
<td>Character Set of Representation</td>
</tr>
<tr>
<td>Accept-Datetime</td>
<td>Memento-Datetime</td>
<td>Time of the Representation</td>
</tr>
</tbody>
</table>

RFC 7089 Defines Memento, Allowing Us To Negotiate In Time
Memento Works In Several Steps

1. **CLIENT:**
   - HEAD <URI-R>
   - Accept-Datetime: DDDDD

2. **CLIENT:**
   - GET <URI-G>
   - Accept-Datetime: DDDDD

3. **CLIENT:**
   - GET <URI-M>

**SERVER:**
- HTTP/1.1 200 OK
- Link: <URI-G, rel=timegate>

**SERVER:**
- HTTP/1.1 302 Found
- Location: <URI-M>
- Vary: Accept-Datetime

**SERVER:**
- HTTP/1.1 200 OK
- Memento-Datetime: DDDDD

**Original Resource (URI-R)**
- **Current Page**

**TimeGate (URI-G)**
- **Resource That Redirects To Archived Pages**

**Memento (URI-M)**
- **Archived Page From The Past**

User running Memento Client
At First The Client Hopes To Get A TimeGate URI From the Server, But...

Most Servers Do Not Yet Return This Data

So Memento Clients Default To Known TimeGates

1. **CLIENT:**
   - HEAD <URI-R>
   - Accept-Datetime: DDDDD

2. **CLIENT:**
   - GET <URI-G>
   - Accept-Datetime: DDDDD

3. **CLIENT:**
   - GET <URI-M>

**SERVER:**
- HTTP/1.1 200 OK
- Link: <URI-G, rel=timegate>

**SERVER:**
- HTTP/1.1 302 Found
- Location: <URI-M>
- Vary: Accept-Datetime

**SERVER:**
- HTTP/1.1 200 OK
- Memento-Datetime: DDDDD

Current Page

Resource That Redirects To Archived Pages

Archived Page From The Past
Then, a browser asks a TimeGate for a specific page from a specific datetime.

GET /web/http://www.cnn.com/ HTTP/1.1
Host: web.archive.org
Accept: image/webp,*/*;q=0.8
Accept-Datetime: Wed, 19 Mar 2003 01:25:35 GMT
Accept-Encoding: gzip, deflate, sdch
Accept-Language: en-US, en; q=0.8
Referer: http://www.cnn.com/
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/36.0.1985.143 Safari/537.36

I Want This Page On This Date
And The TimeGate Tells The Browser Where It Can Go

Go To This URI To Get The Best Memento For The Date You Requested

HTTP/1.1 302 Moved Temporarily
Connection: keep-alive
Content-Type: text/html
Date: Fri, 22 Aug 2014 17:27:57 GMT

Then The Browser Gets A Memento Like Any Other Web Page

This Is The Datetime Of This Memento

Here’s Everything I Know About Archives For That Page
Memento Also Provides TimeMaps As A Machine-Readable List Of Mementos

<http://lostpedia.wikia.com/wiki/The_Numbers>; rel="original",
...deletia
Memento For Chrome
Lets Users “Right-Click Into The Past”

Mink Lets Users See All Mementos For A Page

Who Maintains Mementos?
How Are Mementos Acquired?

1. HTTP GET
   http://en.wikipedia.org/wiki/Spoiler_(media)

2. Pull from link heap
   http://en.wikipedia.org/wiki/Spoiler_(media)

3. Save HTML from fetched page

4. Pass HTML from page to parser to search for embedded resources (links, images, etc.)

5. Put discovered link onto link heap
   http://en.wikipedia.org/wiki/Spoiler_Alert_(How_I_Met_Your_Mother)

---

World Wide Web

- Link heap
   ...

- HTML from page
The Internet Archive Provides The Wayback Machine

http://archive.org/web/

423 Billion web pages saved over time.
Using The Wayback Machine
One Can Select A URI And Datetime...

...And See The Memento From The Internet Archive

The Internet Archive Contains Mementos For Wiki Pages


Abraham Lincoln
From Wikipedia, the free encyclopedia.

Abraham Lincoln (February 12, 1809 - April 15, 1865) was the 16th (1861-1865) President of the United States, and the first President from the Republican Party. He is well praised for successfully restoring the federal unity of the nation by defeating the secessionist Confederate States of America and along the way, playing an important role in ending chattel slavery in the United States. However, a number of states’ rights supporters view Lincoln as a tyrant who suspended civil liberties and suppressed the legitimate right to secede for which Lincoln himself had argued in 1848.

Early life

Born on February 12, 1809, in Kentucky, Lincoln moved at a young age to Indiana then later to New Salem, Illinois. He served as a captain in the U.S. Army during the Black Hawk War. He later tried his hand at several business and political ventures. He was highly regarded as a practicing lawyer. Lincoln served four terms in the Illinois State Legislature, was briefly elected to Congress (1846) and had a successful law practice in Illinois both before and after his single term in the House of Representatives. It is commonly held that Lincoln had turbulent mood swings alternating between grandiosity and depression, which greatly moderated after his marriage to Mary Ann Todd in 1842.

Early political career

First elected to the House of Representatives, Lincoln spent most of his time in Washington, DC alone and made a less than spectacular impression on his fellow politicians. During his unsuccessful campaign for the United States Senate against Stephen A. Douglas, it was Lincoln's well-known gift of oratory that brought public support to an otherwise unimpressive candidate. Lincoln debated Douglas in a series of events which represented a national discussion on the issues that were about to split the nation in two. The Lincoln-Douglas debates presaged the Presidential election of 1860, in which Douglas and Lincoln were once again opponents. On November 6, 1860, Lincoln was elected as the 16th President of the United States, the first Republican to hold that office.

http://en.wikipedia.org/wiki/Abraham_Lincoln
Here Is An Example Wiki Article...

http://en.wikipedia.org/wiki/Abraham_Lincoln
Wikis Are Edited By Multiple Authors, Resulting In Many Revisions...

Most Interesting: Wikis Keep EVERY REVISION OF A PAGE!!!

With Every Revision Preserved, One Can Visit Old Revisions Of Pages

For A Wiki, Every Web Archive Memento Can Be Tied To A Wiki Revision
For a Wiki, every Web Archive Memento can be tied to a Wiki Revision.

We also know which revisions were missed...

The Web Archives are sampling the Wiki Revisions...
Consider the timeline of every episode in a series.
Now Consider A Timeline Of Wiki Revisions Created By Fans
Finally Consider A Third Timeline For Mementos Created From Those Revisions
Steiner Noticed That
Events Inspire Wiki Revisions

... or episodes inspire wiki edits.

http://arxiv.org/abs/1303.4702
Bringing Us To The Naïve Spoiler Concept: Revisions After An Episode Contain Spoilers

\[ t_{r_j} \geq t_{e_i} \implies \text{spoiler} \]
\[ t_{r_j} < t_{e_i} \implies \text{safe} \]
Bringing Us To The Naïve Spoiler Concept: Revisions After An Episode Contain Spoilers

\[ t_{r_j} \geq t_{e_i} \implies \text{spoiler} \]

\[ t_{r_j} < t_{e_i} \implies \text{safe} \]
We Are Not The Only Ones Seeking To Avoid Spoilers...

RELATED WORK
So, What Have Others Done About Spoilers?

Hidden Text!!!

Spoiler Notices!!!

Blocking Social Media Posts!

TO EVERYONE WHO WILL NOT RECEIVE SERIES 4 OF DOWNTON ABBEY UNTIL JANUARY 2014, THIS IS A WARNING: THIS WIKI WILL BE UPDATED AND EDITED AS THE EPISODES BROADCAST IN THE UNITED KINGDOM (SEPTEMBER - DECEMBER 2013) PROCEED AT YOUR OWN RISK! THERE WILL BE SPOILERS ABUNDANT ON MOST OF THE PAGES! YOU HAVE BEEN WARNED!

- Will They or Won't They?: Jim and Pam. They did, sometime between seasons 3 and 4.
  - Andy and Erin.
  - Michael and Holly also seemed poised to become this until they began going out a few episodes in (which Michael hilariously lampshades):
    - Michael: "Why are you helping her? You're not even dating." She's my friend... and... ultimately my strategy is to merge this into a relationship... without her even knowing.
    - [Hidden]

  - Andy and Erin have finally gotten together too.
  - Although Erin dumping Andy in season 6.

Related Work
Existing Academic Studies Have Dealt With Social Media

• Separate studies conducted by Johns and the team of Schirra, Sun, and Bently studied two-screen viewing
• Boyd-Graber, Glasgow, and Zajac attempted to use machine learning to find spoilers in social media
• To avoid spoilers fans would avoid, or abandon:
  - Social media
  - Online web pages
  - TV shows

• This results in lost revenue to advertisers!
Our Goals

• Work with wikis, not social media
• Not just warn the user!
• Not hide the data!

• Show the user what existed before the spoiler was revealed, **so the resource is still useful.**
TimeGates Hold The Key To Avoiding Spoilers

TIMEGATE HEURISTICS
TimeGates Can Be Represented By A Function

\[ M = g^h(R, t_a) \]

- \( M \) = memento returned (URI-M)
- \( R \) = original resource (URI-R)
- \( t_a \) = desired datetime

\( h \) = heuristic being used for TimeGate
mindist Is The Most Widely Used Heuristic

Minimum Distance From $t_a$ With No Bounds

$$G_{mindist}(R, t_8) = m_2 @ t_7$$

$$G_{mindist}(R, t_9) = m_3 @ t_{10}$$

TimeGate Heuristics
But mindist Can Lead To Spoilers...

Minimum Distance – no bounds

\[ g_{\text{mindist}}(R, t_8) = m_2@t_7 \]

\[ g_{\text{mindist}}(R, t_9) = m_3@t_{10} \]
**minpast Does Not Lead To Spoilers**

Minimum Distance In The Past Where Upper Bound = $t_a$

\[ G_{\text{minpast}}(R, t_8) = m_2@t_7 \]
minfutr Is The Opposite Of minpast

Minimum Distance In The Future Where Lower Bound = \( t_a \)

\[
G^{minfutr}(R, t_8) = m_3@t_{10}
\]
minfutr Always Leads To Spoilers

Minimum Distance In The Future Where Lower Bound = $t_a$

$$\mathcal{G}_{minfutr}(R, t_8) = m_3@t_{10}$$
Other Heuristics Can Lead To Spoilers

- **minnear** – bounds specified by user/system
- **eqpast** – compare on both sides of $t_a$, pick past if equal, mindist if not
- **eqfutur** – compare on both sides of $t_a$, pick future if equal, mindist if not
- **simpast** – compare on both sides of $t_a$, pick past if similar, mindist if not
- **simfutur** – compare on both sides of $t_a$, pick future if similar, mindist if not
We Compared These Heuristics Based On Performance And Spoiler Avoidance

TABLE 7: Summary of TimeGate Heuristics

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Uses cost metric</th>
<th>Uses Content of Mementos</th>
<th>Potential running time</th>
<th>Reliably avoids spoilers?</th>
</tr>
</thead>
<tbody>
<tr>
<td>mindist</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>minpast</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>yes</td>
</tr>
<tr>
<td>minfutur</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>minnear</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>eqfutr</td>
<td>X</td>
<td>X</td>
<td>$O(n + c)$</td>
<td>no</td>
</tr>
<tr>
<td>eqpast</td>
<td>X</td>
<td>X</td>
<td>$O(n + c)$</td>
<td>no</td>
</tr>
<tr>
<td>simfutr</td>
<td>X</td>
<td>X</td>
<td>$O(f + n)$</td>
<td>no</td>
</tr>
<tr>
<td>simpast</td>
<td>X</td>
<td>X</td>
<td>$O(f + n)$</td>
<td>no</td>
</tr>
</tbody>
</table>
Even If Modified To Default To minpast, These Heuristics Perform More Poorly

**TABLE 7: Summary of TimeGate Heuristics**

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Uses cost metric</th>
<th>Uses Content of Mementos</th>
<th>Potential running time</th>
<th>Reliably avoids spoilers?</th>
</tr>
</thead>
<tbody>
<tr>
<td>mindist</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>minpast</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>yes</td>
</tr>
<tr>
<td>minfutr</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>minnear</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>eqfutr</td>
<td>X</td>
<td>X</td>
<td>$O(n + c)$</td>
<td>no</td>
</tr>
<tr>
<td>eqpast</td>
<td>X</td>
<td>X</td>
<td>$O(n + c)$</td>
<td>no</td>
</tr>
<tr>
<td>simfutr</td>
<td>X</td>
<td>X</td>
<td>$O(f + n)$</td>
<td>no</td>
</tr>
<tr>
<td>simpast</td>
<td>X</td>
<td>X</td>
<td>$O(f + n)$</td>
<td>no</td>
</tr>
</tbody>
</table>
minpast Is Best For Avoiding Spoilers

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Uses cost metric</th>
<th>Uses Content of Mementos</th>
<th>Potential running time</th>
<th>Reliably avoids spoilers?</th>
</tr>
</thead>
<tbody>
<tr>
<td>mindist</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>minpast</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>yes</td>
</tr>
<tr>
<td>minfutr</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>minnear</td>
<td>X</td>
<td></td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>eqfutr</td>
<td>X</td>
<td>X</td>
<td>$O(n + c)$</td>
<td>no</td>
</tr>
<tr>
<td>eqpast</td>
<td>X</td>
<td>X</td>
<td>$O(n + c)$</td>
<td>no</td>
</tr>
<tr>
<td>simfutr</td>
<td>X</td>
<td>X</td>
<td>$O(f + n)$</td>
<td>no</td>
</tr>
<tr>
<td>simpast</td>
<td>X</td>
<td>X</td>
<td>$O(f + n)$</td>
<td>no</td>
</tr>
</tbody>
</table>
Using mindist Can Be Hazardous For Avoiding Spoilers

THEORY OF SPOILER PROBABILITY
Remember: Consider The Timeline Of Every Episode In A Series
Remember: Consider A Timeline Of Wiki Revisions Created By Fans
Remember: Consider A Third Timeline For Mementos Created From Those Revisions
Remember Our Naïve Spoiler Concept

 Alive

 Deceased

 $t_{r_j} \geq t_{e_i} \implies \text{spoiler}$

 $t_{r_j} < t_{e_i} \implies \text{safe}$
Spoiler Areas Exist Where mindist Returns The User A Memento From The Future Of Their Desired Datetime

In this example, using mindist, if the user requests a memento with a datetime between \( t_9 \) and \( t_{11} \), denoted by the red area, they will get \( m_k \) (which is \( r_j \)), which exists in the future, even though they chose a datetime before \( m_k \)!
Conditions Arising From These Timelines Are Defined By When They Occur

- Pre-Archive – occurs prior to first memento
- Archive-Extant – occurs after first memento
- Post-Archive – occurs after last memento
Condition: Pre Archive Spoiler Area

Spoiler area exists for $e_3$, and one still exists for $e_2$ because $m_1$ maps to $r_j$, which is after $e_3$

\[
[t_s, t_f] = S_a(e_i) = \begin{cases} 
(t_{e_1}, t_{e_i}) & \text{if } t_{e_i} < t_{r_j} \land r_j \equiv m_k \\
(0, 0) & \text{otherwise}
\end{cases}
\]
No spoiler area for $e_3$, but one still exists for $e_2$ because $m_1$ maps to $r_j$, but $r_j$ is before $e_3$
Condition: Archive Extant Spoiler Area

\[ [t_s, t_f] = S_b(e) = \begin{cases} 
(t_h, t_e) & \text{if } t_h < t_e < t_{r_i} \land r_j = m_k \land \frac{t_{m_{k-1}} + t_{m_k}}{2} \\
(0, 0) & \text{otherwise}
\end{cases} \]
Condition: Archive Extant Safe Area

EHR

Theory of Spoiler Probability

\[ [t_s, t_f] = S_b(e) = \begin{cases} (t_h, t_e) & \text{if } t_h < t_e < t_{r_i} \land r_i \equiv m_k \land t_h = \frac{m_{k-1} + m_k}{2} \\ (0, 0) & \text{otherwise} \end{cases} \]
Condition: Archive Extant Safe Area

HRE

\[ [t_s, t_f] = S_b(e) = \begin{cases} (t_h, t_e) & \text{if } t_h < t_e < t_{r_i} \land r_j \equiv m_k \land t_h = \frac{m_{k-1} + m_k}{2} \\ (0, 0) & \text{otherwise} \end{cases} \]
The Area Between the First and Last Episodes Is A Potential Spoiler Zone

Theory of Spoiler Probability
Using Spoiler Areas And A Potential Spoiler Zone, We Can Calculate The Probability Of Spoiler For A Page

\[ Pr(\text{spoiler}) = \frac{s}{c} \]

- \( s \) = # of seconds where we are in a spoiler area
- \( c \) = # of seconds between \( e_1 \) and \( e_n \)
Actual Spoiler Areas From Actual Wikis

MEASUREMENTS OF SPOILER PROBABILITY
Sixteen Lucky Fan Wikis Were Selected From wikia.com...

TABLE 9: Fan wikis used in the spoiler areas experiment

<table>
<thead>
<tr>
<th>Television Show (Network)</th>
<th>Wiki URI .wikia.com</th>
<th># of Pages</th>
<th>$t_{r_1}$</th>
<th>$t_{c_1}$</th>
<th>% of pages in Internet Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>the Big Bang Theory (CBS)</td>
<td>bigbangtheory</td>
<td>1120</td>
<td>2007-12-14</td>
<td>2007-09-24</td>
<td>68.8%</td>
</tr>
<tr>
<td>Boardwalk Empire (HBO)</td>
<td>boardwalkempire</td>
<td>2091</td>
<td>2010-03-18</td>
<td>2010-08-23</td>
<td>80.6%</td>
</tr>
<tr>
<td>Breaking Bad (A&amp;E)</td>
<td>breakingbad</td>
<td>998</td>
<td>2009-04-27</td>
<td>2008-01-20</td>
<td>76.0%</td>
</tr>
<tr>
<td>Continuum (Showcase)</td>
<td>continuum</td>
<td>258</td>
<td>2012-11-13</td>
<td>2012-05-27</td>
<td>86.8%</td>
</tr>
<tr>
<td>Downton Abbey (BBC)</td>
<td>downtonabbey</td>
<td>784</td>
<td>2010-10-04</td>
<td>2010-09-26</td>
<td>53.1%</td>
</tr>
<tr>
<td>Game of Thrones (HBO)</td>
<td>gameofthrones</td>
<td>3144</td>
<td>2010-06-24</td>
<td>2011-04-17</td>
<td>75.8%</td>
</tr>
<tr>
<td>Grimm (NBC)</td>
<td>grimm</td>
<td>1581</td>
<td>2010-04-14</td>
<td>2011-10-28</td>
<td>57.5%</td>
</tr>
<tr>
<td>House of Cards (Netflix)</td>
<td>house-of-cards</td>
<td>251</td>
<td>2013-01-11</td>
<td>2013-02-01</td>
<td>97.2%</td>
</tr>
<tr>
<td>How I Met Your Mother (CBS)</td>
<td>how-i-met-your-mother</td>
<td>1709</td>
<td>2008-07-21</td>
<td>2005-09-19</td>
<td>58.7%</td>
</tr>
<tr>
<td>Lost (ABC)</td>
<td>lostpedia</td>
<td>18790</td>
<td>2005-09-22</td>
<td>2004-09-22</td>
<td>39.1%</td>
</tr>
<tr>
<td>Mad Men (AMC)</td>
<td>madmen</td>
<td>652</td>
<td>2009-07-25</td>
<td>2007-06-03</td>
<td>85.0%</td>
</tr>
<tr>
<td>NCIS (CBS)</td>
<td>ncis</td>
<td>5345</td>
<td>2006-09-25</td>
<td>2003-09-23</td>
<td>93.2%</td>
</tr>
<tr>
<td>Once Upon A Time (ABC)</td>
<td>onceuponatime</td>
<td>1470</td>
<td>2011-08-09</td>
<td>2011-10-23</td>
<td>79.9%</td>
</tr>
<tr>
<td>Scandal (ABC)</td>
<td>scandal</td>
<td>331</td>
<td>2011-06-07</td>
<td>2012-04-05</td>
<td>82.8%</td>
</tr>
<tr>
<td>True Blood (HBO)</td>
<td>trueblood</td>
<td>1838</td>
<td>2008-10-06</td>
<td>2008-09-07</td>
<td>74.1%</td>
</tr>
<tr>
<td>White Collar (USA)</td>
<td>whitecollar</td>
<td>506</td>
<td>2009-10-30</td>
<td>2009-10-23</td>
<td>79.1%</td>
</tr>
</tbody>
</table>
None of these are 100%

In fact, 38% of the total is not available in the Internet Archive.

We See That Not All Pages Are Archived

TABLE 9: Fan wikis used in the spoiler areas experiment

<table>
<thead>
<tr>
<th>Television Show (Network)</th>
<th>Wiki URI .wikia.com</th>
<th># of Pages</th>
<th>tr</th>
<th>te</th>
<th>% of pages in Internet Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>the Big Bang Theory (CBS)</td>
<td>bigbangtheory</td>
<td>1120</td>
<td>2007-12-14</td>
<td>2007-09-24</td>
<td>68.8%</td>
</tr>
<tr>
<td>Boardwalk Empire (HBO)</td>
<td>boardwalkempire</td>
<td>2091</td>
<td>2010-03-18</td>
<td>2010-08-23</td>
<td>80.6%</td>
</tr>
<tr>
<td>Breaking Bad (A&amp;E)</td>
<td>breakingbad</td>
<td>998</td>
<td>2009-04-27</td>
<td>2008-01-20</td>
<td>76.0%</td>
</tr>
<tr>
<td>Continuum (Showcase)</td>
<td>continuum</td>
<td>258</td>
<td>2012-11-13</td>
<td>2012-05-27</td>
<td>86.8%</td>
</tr>
<tr>
<td>Downton Abbey (BBC)</td>
<td>downtonabbey</td>
<td>784</td>
<td>2010-10-04</td>
<td>2010-09-26</td>
<td>53.1%</td>
</tr>
<tr>
<td>Game of Thrones (HBO)</td>
<td>gameofthrones</td>
<td>3144</td>
<td>2010-06-24</td>
<td>2011-04-17</td>
<td>75.8%</td>
</tr>
<tr>
<td>Grimm (NBC)</td>
<td>grrrm</td>
<td>1581</td>
<td>2010-04-14</td>
<td>2011-10-28</td>
<td>57.5%</td>
</tr>
<tr>
<td>House of Cards (Netflix)</td>
<td>house-of-cards</td>
<td>251</td>
<td>2013-01-11</td>
<td>2013-02-01</td>
<td>97.2%</td>
</tr>
<tr>
<td>How I Met Your Mother (CBS)</td>
<td>how-i-met-your-mother</td>
<td>1709</td>
<td>2008-07-21</td>
<td>2005-09-19</td>
<td>58.7%</td>
</tr>
<tr>
<td>Lost (ABC)</td>
<td>lostpedia</td>
<td>18790</td>
<td>2005-09-22</td>
<td>2004-09-22</td>
<td>39.1%</td>
</tr>
<tr>
<td>Mad Men (AMC)</td>
<td>madmen</td>
<td>652</td>
<td>2009-07-25</td>
<td>2007-06-03</td>
<td>85.0%</td>
</tr>
<tr>
<td>NCIS (CBS)</td>
<td>ncis</td>
<td>5345</td>
<td>2006-09-25</td>
<td>2003-09-23</td>
<td>93.2%</td>
</tr>
<tr>
<td>Once Upon A Time (ABC)</td>
<td>onceuponatime</td>
<td>1470</td>
<td>2011-08-09</td>
<td>2011-10-23</td>
<td>79.9%</td>
</tr>
<tr>
<td>Scandal (ABC)</td>
<td>scandal</td>
<td>331</td>
<td>2011-06-07</td>
<td>2012-04-05</td>
<td>82.8%</td>
</tr>
<tr>
<td>True Blood (HBO)</td>
<td>trueblood</td>
<td>1838</td>
<td>2008-10-06</td>
<td>2008-09-07</td>
<td>74.1%</td>
</tr>
<tr>
<td>White Collar (USA)</td>
<td>whitecollar</td>
<td>506</td>
<td>2009-10-30</td>
<td>2009-10-23</td>
<td>79.1%</td>
</tr>
</tbody>
</table>
For Pages With Mementos We Sought To Find Spoiler Areas

Revisions from Wiki export pages

Episode Dates From epguides.com

Mementos from Internet Archive TimeMaps

Pre-Archive Equation

Archive-Extant Equation

List of Spoiler Areas

\[
[t_s, t_f] = S_a(e_i) = \begin{cases} 
(t_{e_1}, t_{e_i}) & \text{if } t_{a_i} < t_{e_j} \land r_j \equiv m_k \\
(0,0) & \text{otherwise}
\end{cases}
\]

\[
[t_s, t_f] = S_b(e) = \begin{cases} 
(t_h, t_e) & \text{if } t_h < t_e < t_{r_i} \land r_j \equiv m_k \land t_h = \frac{t_{m_k} - t_{f_{m_k}}}{2} \\
(0,0) & \text{otherwise}
\end{cases}
\]

Measurements of Spoiler Probability
Spoiler Areas Do Exist!

Spoiler Areas for http://bigbangtheory.wikia.com/wiki/Sheldon_Cooper

Measurements of Spoiler Probability
Spoiler Areas Do Exist!

One Can See Whole Seasons Clumped Together

Fewer Mementos Result In Fewer Chances of Spoiler
We Discovered Different Categories of Pages

- Normal
- Wiki-Before-Show
- Season-In-A-Day
Normal Pages
Are Started After The Series Starts

Spoiler Areas for http://lostpedia.wikia.com/wiki/Kate_Austen
Normal Pages
Are Started After The Series Starts

Pre-Archive
Spoiler Areas
Begin At First Episode

Spoiler Areas for http://lostpedia.wikia.com/wiki/Kate_Austen

Measurements of Spoiler Probability
Wiki-Before-Show Pages Are Started Prior To The First Episode

Spoiler Areas for gameofthrones.wikia.com/wiki/Daenerys_Targaryen

Measurements of Spoiler Probability
Season-in-a-Day Pages Have Many Episodes In A Single Day

Spoiler Areas for http://house-of-cards.wikia.com/wiki/Frank_Underwood
For *Normal* Pages, We See A Variety of Probabilities

**TABLE 11: Spoiler probabilities for most popular pages within each fan wiki**

<table>
<thead>
<tr>
<th>Wiki</th>
<th>Page Name</th>
<th>Probability of Spoiler</th>
<th># of Spoiler Areas</th>
<th># of Revisions</th>
<th># of Mementos</th>
</tr>
</thead>
<tbody>
<tr>
<td>bigbangtheory</td>
<td>Sheldon Cooper</td>
<td>0.31</td>
<td>69</td>
<td>1958</td>
<td>30</td>
</tr>
<tr>
<td>boardwalkempire</td>
<td>Nucky Thompson</td>
<td>0.15</td>
<td>31</td>
<td>290</td>
<td>15</td>
</tr>
<tr>
<td>breakingbad</td>
<td>Walter White</td>
<td>0.43</td>
<td>40</td>
<td>882</td>
<td>20</td>
</tr>
<tr>
<td>continuum</td>
<td>Keira Cameron</td>
<td>0.54</td>
<td>21</td>
<td>104</td>
<td>5</td>
</tr>
<tr>
<td>downtonabbey</td>
<td>Sybil Branson</td>
<td>0.42</td>
<td>23</td>
<td>580</td>
<td>3</td>
</tr>
<tr>
<td>gameofthrones</td>
<td>Daenerys Targaryen</td>
<td>0.16</td>
<td>24</td>
<td>768</td>
<td>29</td>
</tr>
<tr>
<td>grimm</td>
<td>Nick Burkhardt</td>
<td>0.39</td>
<td>30</td>
<td>795</td>
<td>5</td>
</tr>
<tr>
<td>house-of-cards</td>
<td>Frank Underwood</td>
<td>0.00</td>
<td>13</td>
<td>380</td>
<td>3</td>
</tr>
<tr>
<td>how-i-met-your-mother</td>
<td>Barney Stinson</td>
<td>0.55</td>
<td>120</td>
<td>588</td>
<td>13</td>
</tr>
<tr>
<td>lostpedia</td>
<td>Kate Austen</td>
<td>0.67</td>
<td>94</td>
<td>3531</td>
<td>27</td>
</tr>
<tr>
<td>madmen</td>
<td>Mad Men Wiki</td>
<td>0.22</td>
<td>36</td>
<td>250</td>
<td>85</td>
</tr>
<tr>
<td>ncis</td>
<td>Abigail Sciuto</td>
<td>0.67</td>
<td>182</td>
<td>404</td>
<td>11</td>
</tr>
<tr>
<td>onceuponatime</td>
<td>Emma Swan</td>
<td>0.36</td>
<td>34</td>
<td>1210</td>
<td>11</td>
</tr>
<tr>
<td>scandal</td>
<td>Main Page</td>
<td>0.60</td>
<td>31</td>
<td>250</td>
<td>14</td>
</tr>
<tr>
<td>trueblood</td>
<td>Eric Northman</td>
<td>0.28</td>
<td>47</td>
<td>931</td>
<td>14</td>
</tr>
<tr>
<td>whitecollar</td>
<td>Neal Caffrey</td>
<td>0.29</td>
<td>38</td>
<td>199</td>
<td>8</td>
</tr>
</tbody>
</table>
TABLE 11: Spoiler probabilities for most popular pages within each fan wiki

<table>
<thead>
<tr>
<th>Wiki</th>
<th>Page Name</th>
<th>Probability of Spoiler</th>
<th># of Spoiler Areas</th>
<th># of Revisions</th>
<th># of Mementos</th>
</tr>
</thead>
<tbody>
<tr>
<td>bigbangtheory</td>
<td>Sheldon Cooper</td>
<td>0.31</td>
<td>69</td>
<td>1958</td>
<td>30</td>
</tr>
<tr>
<td>boardwalkempire</td>
<td>Nucky Thompson</td>
<td>0.15</td>
<td>31</td>
<td>290</td>
<td>15</td>
</tr>
<tr>
<td>breakingbad</td>
<td>Walter White</td>
<td>0.43</td>
<td>40</td>
<td>882</td>
<td>20</td>
</tr>
<tr>
<td>continuum</td>
<td>Keira Cameron</td>
<td>0.54</td>
<td>21</td>
<td>104</td>
<td>5</td>
</tr>
<tr>
<td>downtonabbey</td>
<td>Sybil Branson</td>
<td>0.42</td>
<td>23</td>
<td>580</td>
<td>3</td>
</tr>
<tr>
<td>gameofthrones</td>
<td>Daenerys Targaryen</td>
<td>0.16</td>
<td>24</td>
<td>768</td>
<td>29</td>
</tr>
<tr>
<td>grimm</td>
<td>Nick Burkhardt</td>
<td>0.39</td>
<td>30</td>
<td>795</td>
<td>5</td>
</tr>
<tr>
<td>house-of-cards</td>
<td>Frank Underwood</td>
<td>0.0</td>
<td>13</td>
<td>380</td>
<td>3</td>
</tr>
<tr>
<td>how-i-met-your-mother</td>
<td>Barney Stinson</td>
<td>0.55</td>
<td>120</td>
<td>588</td>
<td>13</td>
</tr>
<tr>
<td>lostpedia</td>
<td>Kate Austen</td>
<td>0.67</td>
<td>94</td>
<td>3531</td>
<td>27</td>
</tr>
<tr>
<td>madmen</td>
<td>Mad Men Wiki</td>
<td>0.22</td>
<td>36</td>
<td>250</td>
<td>85</td>
</tr>
<tr>
<td>ncis</td>
<td>Abigail Sciuto</td>
<td>0.67</td>
<td>182</td>
<td>404</td>
<td>11</td>
</tr>
<tr>
<td>onceuponatime</td>
<td>Emma Swan</td>
<td>0.36</td>
<td>34</td>
<td>1210</td>
<td>11</td>
</tr>
<tr>
<td>scandal</td>
<td>Main Page</td>
<td>0.60</td>
<td>31</td>
<td>250</td>
<td>14</td>
</tr>
<tr>
<td>trueblood</td>
<td>Eric Northman</td>
<td>0.28</td>
<td>47</td>
<td>931</td>
<td>14</td>
</tr>
<tr>
<td>whitecollar</td>
<td>Neal Caffrey</td>
<td>0.29</td>
<td>38</td>
<td>199</td>
<td>8</td>
</tr>
</tbody>
</table>
Our Model Breaks Down For *Season-In-A-Day*

TABLE 11: Spoiler probabilities for most popular pages within each fan wiki

<table>
<thead>
<tr>
<th>Wiki</th>
<th>Page Name</th>
<th>Probability of Spoiler</th>
<th># of Spoiler Areas</th>
<th># of Revisions</th>
<th># of Mementos</th>
</tr>
</thead>
<tbody>
<tr>
<td>bigbangtheory</td>
<td>Sheldon Cooper</td>
<td>0.31</td>
<td>69</td>
<td>1958</td>
<td>30</td>
</tr>
<tr>
<td>boardwalkempire</td>
<td>Nucky Thompson</td>
<td>0.15</td>
<td>31</td>
<td>290</td>
<td>15</td>
</tr>
<tr>
<td>breakingbad</td>
<td>Walter White</td>
<td>0.43</td>
<td>40</td>
<td>882</td>
<td>20</td>
</tr>
<tr>
<td>continuum</td>
<td>Keira Cameron</td>
<td>0.54</td>
<td>21</td>
<td>104</td>
<td>5</td>
</tr>
<tr>
<td>downtonabbey</td>
<td>Sybil Branson</td>
<td>0.42</td>
<td>23</td>
<td>580</td>
<td>3</td>
</tr>
<tr>
<td>gameofthrones</td>
<td>Daenerys Targaryen</td>
<td>0.16</td>
<td>24</td>
<td>768</td>
<td>29</td>
</tr>
<tr>
<td>grimm</td>
<td>Nick Burkhardt</td>
<td>0.39</td>
<td>30</td>
<td>795</td>
<td>5</td>
</tr>
<tr>
<td>house-of-cards</td>
<td>Frank Underwood</td>
<td>0.0</td>
<td>13</td>
<td>380</td>
<td>3</td>
</tr>
<tr>
<td>how-i-met-your-mother</td>
<td>Barney Stinson</td>
<td>0.55</td>
<td>120</td>
<td>588</td>
<td>13</td>
</tr>
<tr>
<td>lostpedia</td>
<td>Kate Austen</td>
<td>0.67</td>
<td>94</td>
<td>3531</td>
<td>27</td>
</tr>
<tr>
<td>madmen</td>
<td>Mad Men Wiki</td>
<td>0.22</td>
<td>36</td>
<td>250</td>
<td>85</td>
</tr>
<tr>
<td>ncis</td>
<td>Abigail Sciuto</td>
<td>0.67</td>
<td>182</td>
<td>404</td>
<td>11</td>
</tr>
<tr>
<td>onceuponatime</td>
<td>Emma Swan</td>
<td>0.36</td>
<td>34</td>
<td>1210</td>
<td>11</td>
</tr>
<tr>
<td>scandal</td>
<td>Main Page</td>
<td>0.60</td>
<td>31</td>
<td>250</td>
<td>14</td>
</tr>
<tr>
<td>trueblood</td>
<td>Eric Northman</td>
<td>0.28</td>
<td>47</td>
<td>931</td>
<td>14</td>
</tr>
<tr>
<td>whitecollar</td>
<td>Neal Caffrey</td>
<td>0.29</td>
<td>38</td>
<td>199</td>
<td>8</td>
</tr>
</tbody>
</table>

13 spoiler areas exist for this page, but they all have length 0
The Probabilities Do Not Follow A Known Distribution

Histogram of Overall Spoiler Probabilities from 16 Wikia Sites
50% Of The Pages Have A Spoiler Probability < 0.66

Cumulative Distribution Function of Probabilities of Encountering a Spoiler
Using Mindist on Mementos From 16 Wikia Sites

Measurements of Spoiler Probability

99
Remember Missed Updates?

These Revisions Were Missed By The Web Archive...
We Had An Opportunity To Measure Missed Updates

History of Missed Updates Per Day For 15000+ Wiki Pages
We see lines where the archive adopts a more aggressive policy.

Notice how the colors to the right get lighter, indicating fewer missed updates.
Looking At Redundant Mementos, We See The Opposite Effect

Measurements of Spoiler Probability

Notice How The Colors To The Right Get Darker, Indicating More Redundant Mementos
Looking At Redundant Mementos, We See The Opposite Effect

Remember That 38% of the Pages Do Not Have Mementos, So This Does Not Reflect All Archived Pages
SPOILERS IN THE WAYBACK MACHINE

The Wayback Machine Uses mindist, Ergo Users Encounter Spoilers There...
We Have Access To Some Anonymized Logs From The Wayback Machine

From these logs, we have:

- which memento the user came from (the referer)
- the memento returned to the user

The Wayback Machine Rewrites URIs, Redirecting Users When They Click On Links To Other Pages That Exist During The Same Time Period.
We Can Infer Desired Datetime And Have The Memento-datetime

http://web.archive.org/web/20020404020224/http://www.example.com

From the referrer, we can infer their desired datetime

From the memento returned to the user, we can acquire the Memento-Datetime
Using the Logs, We Can See How Many wikia.com Requests End In Spoilers

Wayback Logs

Desired Datetime From Referrer

Memento-datetime From Visited URI

Revisions From Wiki Export Pages

Datetime Of Revision

Did The Wayback Machine Deliver The User To A Revision That Existed In The Future From The Desired Datetime?

No

Safe

Yes

Spoiler

Spoilers In The Wayback Machine
Wayback Machine Log Results For wikia.com Requests

We Have No Way To Identify Wikis Other Than By Domain Name, Data Does Not Include All Wikis...
A Solution For Avoiding Spoilers In Wikis. Wiki Revisions Are Mementos, Too!

THE MEMENTO MEDIAWIKI EXTENSION
We Developed The Memento MediaWiki Extension To Use minpast

http://www.mediawiki.org/wiki/Extension:Memento

MediaWiki Is The Software Used By Wikipedia And Wikia.Com
We Set Up A Demonstration Wiki...

http://ws-dl-05.cs.odu.edu/demo/index.php/Main_Page

Data was exported from: http://awoiaf.westeros.org
For Performance, We Had An Opportunity To Experiment With Alternate Memento Patterns
Original Resource Acts As Own TimeGate (Pattern 1.1)

This Pattern Only Requires 2 Requests To Acquire A Memento. Intuitively, It Should Perform Better.
Client Must Ask About TimeGate (Pattern 2.1)

This Pattern Requires 3 Requests To Acquire A Memento. Intuitively, It Should Perform Worse.
Using Analysis and Experimentation
We Found That The Three-Request Pattern Actually Performed Better

Mediawiki Takes Too Long To Generate A TimeGate Response If The Two Are The Same Page.

The Three-request Pattern Requires An Extra Request, But, Due To Round-trip Time, It Performs Better Unless The User Has A Bandwidth Of 21,926 bps Or Less.
Serving Current Pages Takes The Same Time With Or Without The Extension

We used seige for testing, as detailed in [http://arxiv.org/abs/1406.3876](http://arxiv.org/abs/1406.3876)

We found that:

- **Mean** = -0.0072 s
- **Std dev** = 0.3526 s
Serving Old Pages (Mementos) Takes The Same Time With Or Without The Extension

We used seige for testing, as detailed in http://arxiv.org/abs/1406.3876

Mean = -0.0026 s
Std dev = 0.0421 s
TimeMaps Are Smaller Than Wiki History Pages, So They Perform Better

Differences in size between Memento MediaWiki Extension TimeMaps and MediaWiki History Pages

We used seige for testing, as detailed in [http://arxiv.org/abs/1406.3876](http://arxiv.org/abs/1406.3876)
MediaWiki Does Not Just Store Previous Revisions Of Pages...

Images, Stylesheets, and JavaScript have previous revisions stored as well
Then We Wondered, What About Images? They Could Contain Spoilers...

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Dimensions</th>
<th>User</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>21:46, 22 March</td>
<td>959 x 593</td>
<td>Drafwik</td>
<td>Retract Michigan</td>
</tr>
<tr>
<td>15:55, 22 March</td>
<td>959 x 593</td>
<td>Drafwik</td>
<td>Implement beige stripes per en:File talk:Sameness marriage in USA.svg/Proposal 3 (Change black to beige/gold)</td>
</tr>
<tr>
<td>21:37, 21 March</td>
<td>959 x 593</td>
<td>Drafwik</td>
<td>Michigan struck down, no stay yet</td>
</tr>
<tr>
<td>01:03, 20 March</td>
<td>959 x 593</td>
<td>Drafwik</td>
<td>Moving Kentucky to stayed ruling category</td>
</tr>
<tr>
<td>01:24, 4 March</td>
<td>959 x 593</td>
<td>Drafwik</td>
<td>Implement stripes for states with stayed rulings per consensus at en:File talk:Sameness marriage in USA.svg</td>
</tr>
<tr>
<td>15:36, 1 March</td>
<td>959 x 593</td>
<td>SPORobin</td>
<td>re-add Kentucky recognition, since it is equivalent to a law that has been signed but not taken effect yet</td>
</tr>
<tr>
<td>06:46, 1 March</td>
<td>959 x 593</td>
<td>Plumber</td>
<td>revision until consensus is found</td>
</tr>
<tr>
<td>08:08, 1 March</td>
<td>959 x 593</td>
<td>Plumber</td>
<td>proposed changes to the territories, switching up limited/foreign</td>
</tr>
</tbody>
</table>

The Memento MediaWiki Extension
Consider This Map

This Map is important to understanding the content of this article.

This image is changed as the article is changed, to reflect its content.

http://en.wikipedia.org/wiki/Same-sex_marriage_law_in_the_United_States_by_state
It’s The Same Map If Today We Visit
The June 5, 2013 Revision

Users can't view this embedded resource as it looked on June 2013 while reading the article from that time period.

http://en.wikipedia.org/w/index.php?title=Same-sex_marriage_law_in_the_United_States_by_state&oldid=558400004
What Should Have Happened

This is the current map

This is the map from June, 2013 that should have been displayed

The content of the article won't match the data in this visual aid, possibly confusing a user who wanted historical information on this topic.
We Developed A Solution For Images

The $file argument’s getHistory() function of the ImageBeforeProduceHTML hook can be used to acquire previous revisions of images
We Could Not Extract Previous Revisions Of CSS And Javascript...

The data is present, but we could not find any way for an extension to access or render it.
We Demonstrated Avoiding Spoilers
At WikiConUSA 2014...

We want to find information about Kevan Lannister, but haven’t read the book *A Dance with Dragons* yet. We set the Memento Chrome Extension prior to the release of that book: June 29, 2011.
So We Set Memento For Chrome To The Correct Date...

We use the Memento Chrome Extension to request a revision of the page close to, but not over, our requested date.
...And Got A Page Without Spoilers

And We Avoid Spoilers for A Dance With Dragons...
FUTURE WORK

Where Do We Go From Here?

FUTURE WORK

• Background
• Related Work
• TimeGate Heuristics
• Theory of Spoiler Probability
• Measurements of Spoiler Probability
• Spoilers In The Wayback Machine
• The Memento MediaWiki Extension
• Future Work
• Conclusions
How Do We Handle Spoilers For 
*Season-In-A-Day* Series?
Can We Create A New Heuristic Based On mindist For Spoiler Detection?

Can we process the content of $m_3$ and redirect the user to $m_2$ if we detect spoilers?
Can We Use The Extension To Avoid Spoilers For Sports And The News?
How Do We Use minfutr And minpast To Study Emerging Topics On Wikipedia

Future Work
We Can Do Further Work On Missed Updates And Redundant Mementos
CONCLUSIONS

We Got Here...

CONCLUSIONS
We Introduced The Naïve Spoiler Concept

\[ t_{r_j} \geq t_{e_i} \Rightarrow \text{spoiler} \]

\[ t_{r_j} < t_{e_i} \Rightarrow \text{safe} \]
## We Have Detailed TimeGate Heuristics

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Uses cost metric</th>
<th>Uses Content of Mementos</th>
<th>Potential running time</th>
<th>Reliably avoids spoilers?</th>
</tr>
</thead>
<tbody>
<tr>
<td>mindist</td>
<td>X</td>
<td>X</td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td><strong>minpast</strong></td>
<td><strong>X</strong></td>
<td></td>
<td>$O(n)$</td>
<td><strong>yes</strong></td>
</tr>
<tr>
<td>minfutr</td>
<td>X</td>
<td>X</td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>minnear</td>
<td>X</td>
<td>X</td>
<td>$O(n)$</td>
<td>no</td>
</tr>
<tr>
<td>eqfutr</td>
<td>X</td>
<td>X</td>
<td>$O(n+c)$</td>
<td>no</td>
</tr>
<tr>
<td>eqpast</td>
<td>X</td>
<td>X</td>
<td>$O(n+c)$</td>
<td>no</td>
</tr>
<tr>
<td>simfutr</td>
<td>X</td>
<td>X</td>
<td>$O(f+n)$</td>
<td>no</td>
</tr>
<tr>
<td>simpast</td>
<td>X</td>
<td>X</td>
<td>$O(f+n)$</td>
<td>no</td>
</tr>
</tbody>
</table>
We showed how to calculate the probability of encountering a spoiler.

\[ Pr(\text{spoiler}) = \frac{s}{c} \]

- \( s \) = # of seconds where we are in a spoiler area
- \( c \) = # of seconds between \( e_1 \) and \( e_n \)
We Calculated Real Spoiler Probabilities

TABLE 11: Spoiler probabilities for most popular pages within each fan wiki

<table>
<thead>
<tr>
<th>Wiki</th>
<th>Page Name</th>
<th>Probability of Spoiler</th>
<th># of Spoiler Areas</th>
<th># of Revisions</th>
<th># of Mementos</th>
</tr>
</thead>
<tbody>
<tr>
<td>bigbangtheory</td>
<td>Sheldon Cooper</td>
<td>0.31</td>
<td>69</td>
<td>1958</td>
<td>30</td>
</tr>
<tr>
<td>boardwalkempire</td>
<td>Nucky Thompson</td>
<td>0.15</td>
<td>31</td>
<td>290</td>
<td>15</td>
</tr>
<tr>
<td>breakingbad</td>
<td>Walter White</td>
<td>0.43</td>
<td>40</td>
<td>882</td>
<td>20</td>
</tr>
<tr>
<td>continuum</td>
<td>Keira Cameron</td>
<td>0.54</td>
<td>21</td>
<td>104</td>
<td>5</td>
</tr>
<tr>
<td>downtonabbey</td>
<td>Sybil Branson</td>
<td>0.42</td>
<td>23</td>
<td>580</td>
<td>3</td>
</tr>
<tr>
<td>gameofthrones</td>
<td>Daenerys Targaryen</td>
<td>0.16</td>
<td>24</td>
<td>768</td>
<td>29</td>
</tr>
<tr>
<td>grimm</td>
<td>Nick Burkhardt</td>
<td>0.39</td>
<td>30</td>
<td>795</td>
<td>5</td>
</tr>
<tr>
<td>house-of-cards</td>
<td>Frank Underwood</td>
<td>0.0</td>
<td>13</td>
<td>380</td>
<td>3</td>
</tr>
<tr>
<td>how-i-met-your-mother</td>
<td>Barney Stinson</td>
<td>0.55</td>
<td>120</td>
<td>588</td>
<td>13</td>
</tr>
<tr>
<td>lostpedia</td>
<td>Kate Austen</td>
<td>0.67</td>
<td>94</td>
<td>3531</td>
<td>27</td>
</tr>
<tr>
<td>madmen</td>
<td>Mad Men Wiki</td>
<td>0.22</td>
<td>36</td>
<td>250</td>
<td>85</td>
</tr>
<tr>
<td>ncis</td>
<td>Abigail Sciuto</td>
<td>0.67</td>
<td>182</td>
<td>404</td>
<td>11</td>
</tr>
<tr>
<td>onceuponatime</td>
<td>Emma Swan</td>
<td>0.36</td>
<td>34</td>
<td>1210</td>
<td>11</td>
</tr>
<tr>
<td>scandal</td>
<td>Main Page</td>
<td>0.60</td>
<td>31</td>
<td>250</td>
<td>14</td>
</tr>
<tr>
<td>trueblood</td>
<td>Eric Northman</td>
<td>0.28</td>
<td>47</td>
<td>931</td>
<td>14</td>
</tr>
<tr>
<td>whitecollar</td>
<td>Neal Caffrey</td>
<td>0.29</td>
<td>38</td>
<td>199</td>
<td>8</td>
</tr>
</tbody>
</table>
We showed that the Wayback Machine is serving spoilers.

Wayback Machine Requests For wikia.com:
Spoilers vs Safe

- 81% for safe
- 19% for spoilers
We Developed The Memento MediaWiki Extension To Use minpast
Most of All We Showed That It Is Possible To Avoid Spoilers In MediaWiki Fan Sites Using Memento


Accept-Datetime: Sun, 13 April 2014 00:59:00 GMT
Papers/Presentations

Because What You Have Seen Is Just The Tip Of The Iceberg...

BACKUP SLIDES
Backup Slides

RELATED WORK
Existing Studies: Leavitt and Christenfeld

![Bar chart showing mean hedonic ratings for various stories by Leavitt and Christenfeld. The x-axis categorizes stories into Ironic-Twist Stories, Mysteries, and Literary Stories. The y-axis represents mean hedonic ratings. Bars indicate spoiled and unspoiled conditions for each story type.](chart.png)
Existing Studies: Release Date Different Per Country

• Schirra, Sun, and Bently’s live-tweeting study on *Downton Abbey* unearthed a global problem of avoiding spoilers in shows that air later for fans in a different country.

• This is further echoed by Leaver’s essay *The Tyranny of Digital Distance* discussing the release of *Battlestar Galactica* episodes in Australia.
Existing Studies: Johns

- Johns studied two-screen viewing (generic name for live tweeting)
- Fans of particular shows would avoid social media until they had viewed the latest show
- These fans were trying desperately to avoid spoilers
- This contradicts Leavitt’s study
Typically Notices Warn Us of Spoilers

*The World of Ice and Fire (US, UK)* will be released October 28. Editors, please refrain from including information from the book prior to November 27.

**Synopsis**

*Note: The following synopsis refers to the "Theatrical cut" version of the film, which features departures from the original novel.*

*Spoiler warning! Plot and/or ending details follow.*

**TO EVERYONE WHO WILL NOT RECEIVE SERIES 4 OF DOWNTON ABBEY UNTIL JANUARY 2014, THIS IS A WARNING: THIS WIKI WILL BE UPDATED AND EDITED AS THE EPISODES BROADCAST IN THE UNITED KINGDOM (SEPTEMBER - DECEMBER 2013)!!! PROCEED AT YOUR OWN RISK! THERE WILL BE SPOILERS ABUNDANT ON MOST OF THE PAGES! YOU HAVE BEEN WARNED!**

**MAJOR SPOILER ALERT:** This story contains many details of Thursday’s season finale episode of *ABC’s Scandal.*

**Related Work**
Some Sites Try To Hide The Spoilers With HTML And/OR Javascript

- Will They or Won’t They?: Jim and Pam. They did, sometime between seasons 3 and 4.
  - Andy and Erin.
  - Michael and Holly also seemed poised to become this until they began going out a few episodes in (which Michael hilariously lampshades):
    
    Michael: "Why are you helping her? You’re not even dating. She’s my friend... and... ultimately my strategy is to merge this into a relationship... without her even knowing.

- Andy and Erin have finally gotten together too.
- Although Erin dumps Andy in season 9.
Existing Academic Studies Have Dealt With Social Media

• Separate studies conducted by Johns and the team of Schirra, Sun, and Bently studied two-screen viewing
• Boyd-Graber, Glasgow, and Zajac attempted to use machine learning to find spoilers in social media
• To avoid spoilers fans would avoid, or abandon:
  – Social media
  – Online web pages
  – TV shows
• This results in lost revenue to advertisers!
Existing Studies: Tsang and Yan

• Fans have abandoned
  – Social media
  – Online web pages
  – TV shows

• Because of the issues in avoiding spoilers

• This results in lost revenue to advertisers!
Hiding Spoilers: Microformats

Artjom Kurapov created a draft HTML microformat for classifying links and images.

Listing 3.1: Examples of xrate microformats for avoiding spoilers, pornography, and violence in links and images

```html
<a href="http://www.example.com/who-my-character-fell-in-love-with" data-xrate-spoiler="100" data-xrate-sex="20" >link on information about this episode</a>

<img src="http://www.example.com/picture-of-character-finally-dying" data-xrate-spoiler="100" data-xrate-violence="60" />
```
Hiding Spoilers: Tweetdeck

Related Work
Hiding Spoilers: Tumblr Savior

gameofthronesdaily made a post containing: 'sansa stark' -- click to show.

Tags: #reblog#cast

Related Work
Hiding Spoilers: Netflix Spoiler Foiler
Hiding Spoilers: Netflix Spoiler Foiler
Spoiler Shield Is A Spoiler Blocking Tool For Social Media

<table>
<thead>
<tr>
<th>S</th>
<th>Spoiler Shield</th>
</tr>
</thead>
</table>

**SHIELDS**

- **Television**
- **MLB**
- **NFL**
- **College Football**
- **Celebrities**

**TELEVISION**

- Dancing With The Stars
- Days Of Our Lives
- Downton Abbey
- Dr. Who
- Elementary
- Fargo
- Game Of Thrones
- Girls
- Glee
- Grey's Anatomy
- Hell's Kitchen
- Homeland
- House of Cards

**Related Work**

- "Game Of Thrones" Shield by Entertainment Weekly
  - Share on Facebook
  - Share on Twitter
  - Share by Mail

- "Game Of Thrones" Shield by
  - Share on Facebook
  - Share on Twitter
  - Share by Mail

- Rick Broda
- Edit Profile
- News Feed
- Messages
- Events

- Viewing most recent stories
- Back to top stories

- Lifehacker
- 23 mins - €

**TRENDING**

- Chris Brown
  - Jail, LA deposition
- Apple World Conference: 2014 Apple TV
- Ann B. Davis
  - Brady Bunch See More

**GROUPS**

- Groups You Admin
  - Foxcroft Homeowners
  - North Farmington
  - For sale, buy, rent...
  - Create Group

**APPS**

- Games
- Candy Crush Saga
- toGather
- Cities I've Visited
- MyLowe's Home Makeover
- Spotify
- Roloi 3 Sweeps
- Zuma Blitz
- N.O.V.A. Near Orbit V...
- Lexulous Word Game
Hiding Spoilers: Facebook Posts Filter And Open Tweet Filter

Any Facebook posts with these keywords will not be displayed.

Any tweets with these keywords will not be displayed.
Archiving In MediaWiki: Parsoid

Related Work
Archiving In MediaWiki: Collection Extension

Collection
From Simple English Wikipedia - the free encyclopedia that anyone can change

You can collect articles, generate and download a PDF file from article collections, order books from a print-on-demand partner and save article collections for later use or to share them.
See Help:Collections for more information about collections.

My Collection

Title: Music Wiki Selection
Subtitle: Articles from Wikipedia

Contents
[Create new chapter] [Sort articles alphabetically] [Clear collection]

• Basic topics
  • Falsetto [Remove]
  • Singing [Remove]
  • Soprano [Remove]
  • Mezzo-soprano [Remove]
  • Baritone [Remove]
  • Harpsichord [Remove]
• Composers
  • Ludwig van Beethoven [Remove]
  • Johannes Brahms [Remove]
  • Wolfgang Amadeus Mozart [Remove]
  • Joseph Haydn [Remove]

Order Printed Book
You can order a printed book containing your article collection by visiting one of the following print-on-demand partners:
- Order book from PediaPress - About PediaPress

Download Collection as PDF
To download an automatically generated PDF file of your article collection, click the button.
[Download PDF]

Save Collection
If you want to save collections for later use, please log in or create an account.
Seamlessly Viewing Past Versions Of MediaWiki Pages

1. CLIENT: HEAD <URI-R> Accept-Datetime: DDDDD
2. CLIENT: GET <URI-G> Accept-Datetime: DDDDD
3. SERVER: HTTP/1.1 200 OK Location: <URI-M> Vary: Accept-Datetime
4. SERVER: HTTP/1.1 200 OK with JSON response

Related Work
Seamlessly Viewing Past Versions Of MediaWiki Pages

Wikipedia Proxy Does Not Address The Problem For All Wikis.

To Be Generic For Any Wiki, The Wiki Server Itself Would Need To Send A Link Header Back Indicating The Uri Of The TimeGate To Use.
ALTERNATIVES TO THE MEMENTO MEDIAWIKI EXTENSION (AND WHY IT IS BETTER!)
Memento Extension vs. Manually Getting Page Revision

This Is Very Time Consuming.

Memento Let’s You *Browse* Through The Whole Web With A Given Date!

Why Do It When Memento Will Do It For You?
Memento Extension vs. MediaWiki API

JSON:

XML:
<rev revid="607519915" parentid="607345961" user="Marklemagne" timestamp="2014-05-07T19:00:26Z"/>

Only A Custom MediaWiki Client Can Turn These Oldid Entries Into Uris.

Memento Is A Web Standard Way Of Accessing Past Web Resources And Is Already Implemented For Many Different Applications (Web Archives, Etc.)
Memento Extension vs. MediaWiki API

Memento Also Follows The RESTful Principle Of “Follow Your Nose”, Indicating Additional Resources To Access From Here.
Memento Extension vs. Internet Archive

The Internet Archive Only Gets Some Of The Revisions Of A Given Page.

MediaWiki Has All Of The Revisions Of A Given Page.
Memento Extension vs. Other MediaWiki Time Travel Extensions

While These Extensions Just Work For MediaWiki, **Memento Works For The Entire Web.**

With The Memento Extensions, One Can Browse The Entire Web Spoiler Free By Seamlessly Accessing Web Archives And Other Resources Through Memento.
Backup Slides

MEMENTO PATTERN COMPARISON IN THE MEMENTO MEDIAWIKI EXTENSION
Assuming An Original Resource Is a TimeGate (Pattern 1.1)

\[ d_{p1.1} = B + RTT_B + M + RTT_M \]
Looking For A TimeGate (Pattern 2.1)

\[ d_{p2.1} = a + RTT_a + b + RTT_b + M + RTT_M \]
Comparing These Two Involves Evaluating Their Performance

\[ \frac{d_{p1.1}}{d_{p1.2}} \]

\[ B + RTT_B + \mathcal{M} + RTT_M < a + RTT_a + b + RTT_b + \mathcal{M} + RTT_M \]

\[ B + RTT_B < a + RTT_a + b + RTT_b \]

\[ B < a + RTT_a + b \]

\[ B < a + b + RTT_a \]

\( a \) = time to generate just a normal wiki page
\( b \) = time to perform datetime negotiation when the TimeGate is the same
\( B \) = time to perform datetime negotiation when the TimeGate is different
\( M \) = time to generate memento
\( RTT_{a}, RTT_{b}, RTT_{B}, RTT_{M} \) = round trip times for \( a, b, B, \) and \( M \)
Using curl
We Obtained the Value of $a$

$$B < a + b + RTT_a$$

With caching, $a$ turns out to be about 0.1 seconds on average
Using seige
We Obtained Values for $b$ and $B$

$B < a + b + RTT_a$

From these results:

$0.22 \leq b \leq 1.75$

$0.56 \leq B \leq 12.06$
Comparing TimeGate Implementations

Differences in URI-G Performance
For the Memento Mediawiki Extension
Between URI-R=URI-G and SpecialPage TimeGate Implementations

Worse Performance
Using Analysis We Worked To Obtain The Value Of $RTT_a$

Round Trip Time is the sum of transmission delay and propagation delay

$$RTT = d_t + d_p$$

Transmission delay is # of bits divided by the rate of transmission

$$d_t = \frac{N}{R}$$

Requests and responses are typically about 11,840 bits

Assuming a worst case of 1G telephony (28,000 bps), $d_t = 0.41 \text{ s}$

Using our previous values for $a$, $b$, and $B$, we see that such a 1G user would need to experience a $d_p$ of 0.13s for the two-request pattern to perform better.
Continuing Our Analysis We Obtained The Value Of $RTT_a$

0.13s sounds small, but at the speed of light, this would require the user to be 24,216.7 miles from the server

This is almost the circumference of the Earth!!!
So, At What Value Of $d_t$ Does The Two-Request Pattern Win Out?

\[
1.24 \, s < d_t + 0.1 \, s + 0.6 \, s
\]
\[
1.24 \, s < d_t + 0.7 \, s
\]
\[
0.54 \, s < d_t
\]
\[
d_t = \frac{N}{R} \quad \text{From (20)}
\]
\[
0.54 \, s = \frac{11840 \, b}{R}
\]
\[
(0.54 \, s)(R) = 11840 \, b
\]
\[
R = \frac{11840 \, b}{0.54 \, s} = 21926 \, bps
\]

At bandwidth less than 21,926 bps, the two-request pattern wins out.

That’s slightly better than 1G telephony!

For most users, the three-request pattern performs better!
Backup Slides

ADDITIONAL ARCHIVE EXTANT SAFE AREAS
These Are Identified By The Order Of Occurrence Of Halfway, Revision, and Event

• For example
  – HRE – means:
    1. Halfway mark between two mementos
    2. Revision is created
    3. Event occurs
  – ERH – means:
    1. Event occurs
    2. Revision is created
    3. Halfway mark between two mementos
Condition: Archive Extant Safe Area

REH

\[
[t_s, t_f] = S_b(e) = \begin{cases} 
(t_h, t_e) \quad & \text{if} \quad t_h < t_e < t_{r_i} \land r_j \equiv m_k \land \ \frac{t_{m_{k-1}} + t_{m_k}}{2} \\
(0, 0) \quad & \text{otherwise}
\end{cases}
\]
Condition: Archive Extant Safe Area

RHE

\[
[t_s, t_f] = S_b(e) = \begin{cases} 
(t_h, t_e) & \text{if } t_h < t_e < t_{r_i} \land r_j \equiv m_k \land \\
(0, 0) & \text{otherwise}
\end{cases}
\]

Theory of Spoiler Probability
Condition: Archive Extant Safe Area

ERH

\[ [t_s, t_f] = S_b(e) = \begin{cases} 
(t_h, t_e) & \text{if } t_h < t_e < t_{r_i} \land r_j \equiv m_k \land t_h = \frac{t_{m_{k-1}} + t_{m_k}}{2} \\
(0, 0) & \text{otherwise} 
\end{cases} \]
Condition: Post Archive Safe Area RE

No spoilers after last memento $m_n$
Condition: Post Archive Safe Area ER

No spoilers after last memento $m_n$
ARCHITECTURE OF THE MEMENTO MEDIAWIKI EXTENSION
We Used Class Inheritance For the Different Memento Resource Types

http://www.mediawiki.org/wiki/Extension:Memento
MediaWiki SpecialPages Invoke TimeGate And TimeMap Functionality

http://www.mediawiki.org/wiki/Extension:Memento
All Datetime Negotiation Is Centralized In The TimeNegotiator Class

http://www.mediawiki.org/wiki/Extension:Memento
The Memento Class Is The Entry Point For The Extension

http://www.mediawiki.org/wiki/Extension:Memento
Backup Slides

MISCELLANEOUS
MediaWiki Still Has CSS Issues
Other Uses For The Memento MediaWiki Extension

Evolving laws and legal discourse

Changing relationship between organizations (ICANN vs. Verisign)

Past software contributions (Folding@Home)
Memento Headers Extension

1. **CLIENT:**
   - HEAD <URI-R>
   - Accept-Datetime: DDDDD

2. **CLIENT:**
   - GET <URI-G>
   - Accept-Datetime: DDDDD

3. **CLIENT:**
   - GET <URI-M>

**SERVER:**
- HTTP/1.1 200 OK
- Link: <URI-G, rel=timegate>

**SERVER:**
- HTTP/1.1 302 Found
- Location: <URI-M>
- Vary: Accept-Datetime

**Original Resource (URI-R) → TimeGate (URI-G) → Memento MediaWiki Proxy TimeGate → Memento (URI-M) → Wiki running Memento Headers MediaWiki Extension**

User running Memento Client