Human-Computer Interaction Lab
University of Maryland

LifeFlow Case Study:
How do people read children books online?

This case study was conducted by Krist Wongsuphasawat and Hsueh-Chien Cheng
under the supervision of Ben Shneiderman and Catherine Plaisant.
We appreciated the feedback from Anne Rose of the ICDL.

Last updated: Nov 22, 2011

http://www.cs.umd.edu/hcil/lifeflow

Introduction

We looked at the Apache web logs from the International Children's Digital Library to analyze how people read children books online. This sample dataset was taken from the period Jul 01-07, 2011. The original size of the dataset was about 1GB.

We filtered the log entries to select only the http requests to access an actual book page:

```
"http://www.childrenslibrary.org/icdl/BookPage"
```

Each request contains a query string behind the url, such as

```
"bookid=amrdima_00310002&pnum1=6&pnum2=7&twoPage=true&lang=English&ilang=English"
```

We parsed the page number and book id from this query string then group the http requests by IP address and book id into records. Each record represents how each IP address (user) read one book, or a *book session*.

```
IP address + book id            Page    Time
133.37.60.191_radjese_00380046  2       2011-06-30 05:01:06
133.37.60.191_radjese_00380046  1       2011-06-30 05:01:14
133.37.60.191_radjese_00380046  2       2011-06-30 05:01:20
133.37.60.191_radjese_00380046  4       2011-06-30 05:01:25
133.37.60.191_radjese_00380046  6       2011-06-30 05:01:27
```

We included only the first 30 pages in the dataset. The processed dataset has 7,022 records and 57,709 rows in total.

LifeFlow 3.0 - untitled (bookpage_byIPAndBook_icdl0711.txt)

File Dataset Selection View Help

Data Manager | Control

untitled (bookpage_byIPAnd...

showing 7022/7022 records

0 selected

Remove them | Remove others

All | Overview | Timeline

Align | Filter

Please specify how to align:

No alignment

1 st ▲ 1 (3024)

from the beginning

All ▲ 1 (3024)

(no advanced option selected)

Advanced Options...

automatic update | Align

Legend | Attribute

- ▲ 1
- ▲ 2
- ▲ 4
- ▲ 6
- ▲ 8
- ▲ 10
- ▲ 12
- ▲ 14
- ▲ 16

Measurement Tool

Zoom: 12 Days

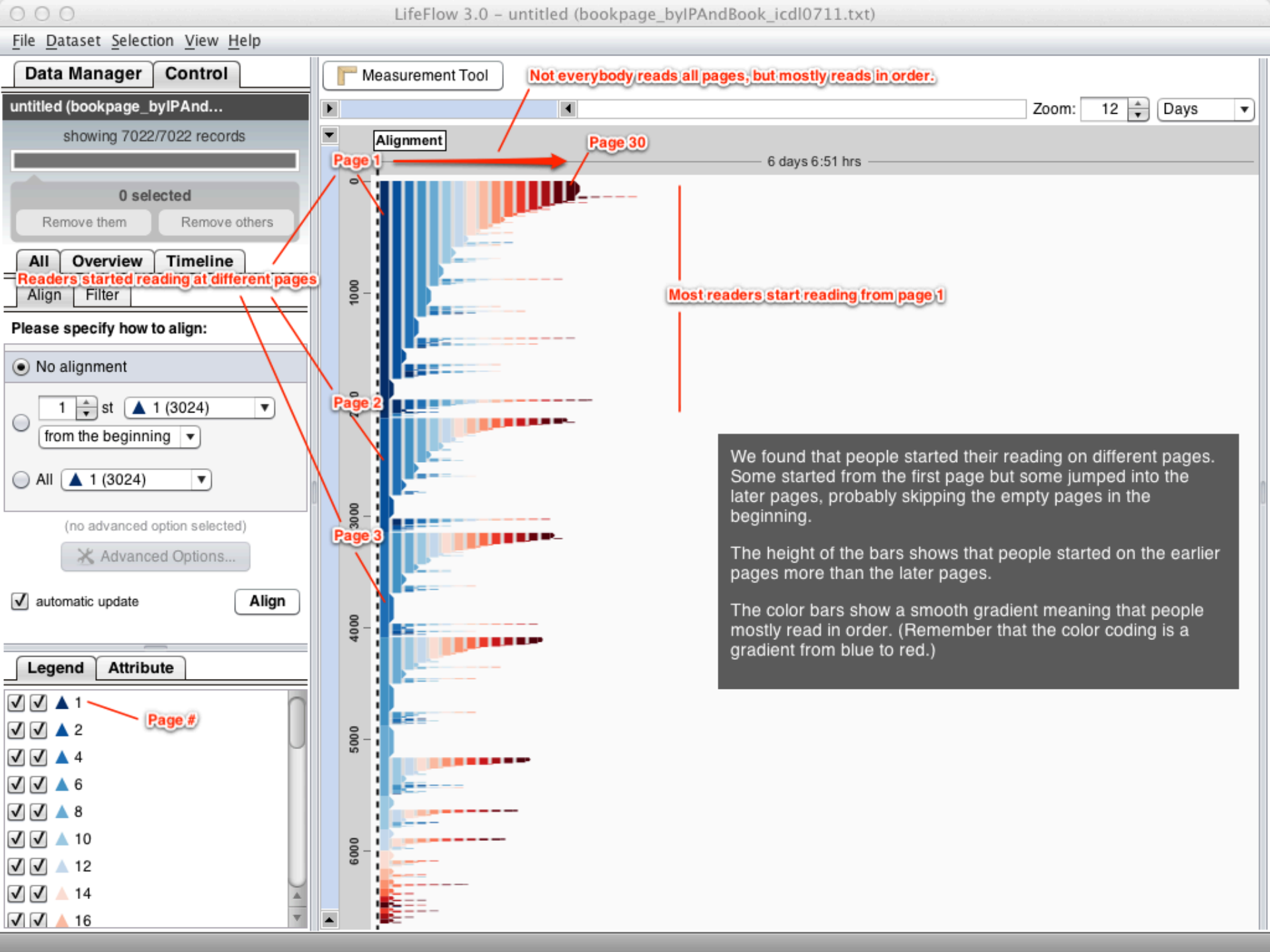
Alignment

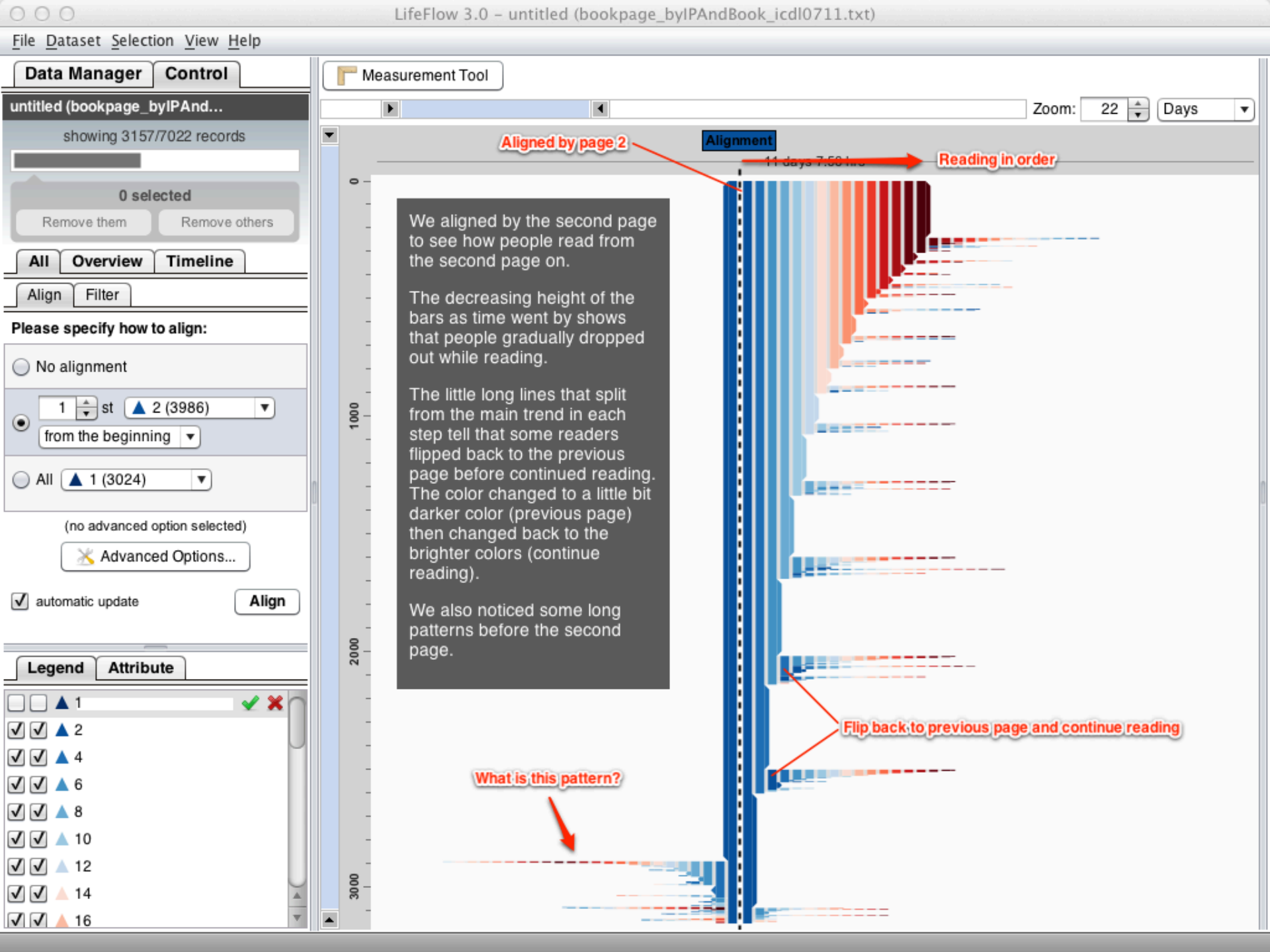
6 days 6:51 hrs

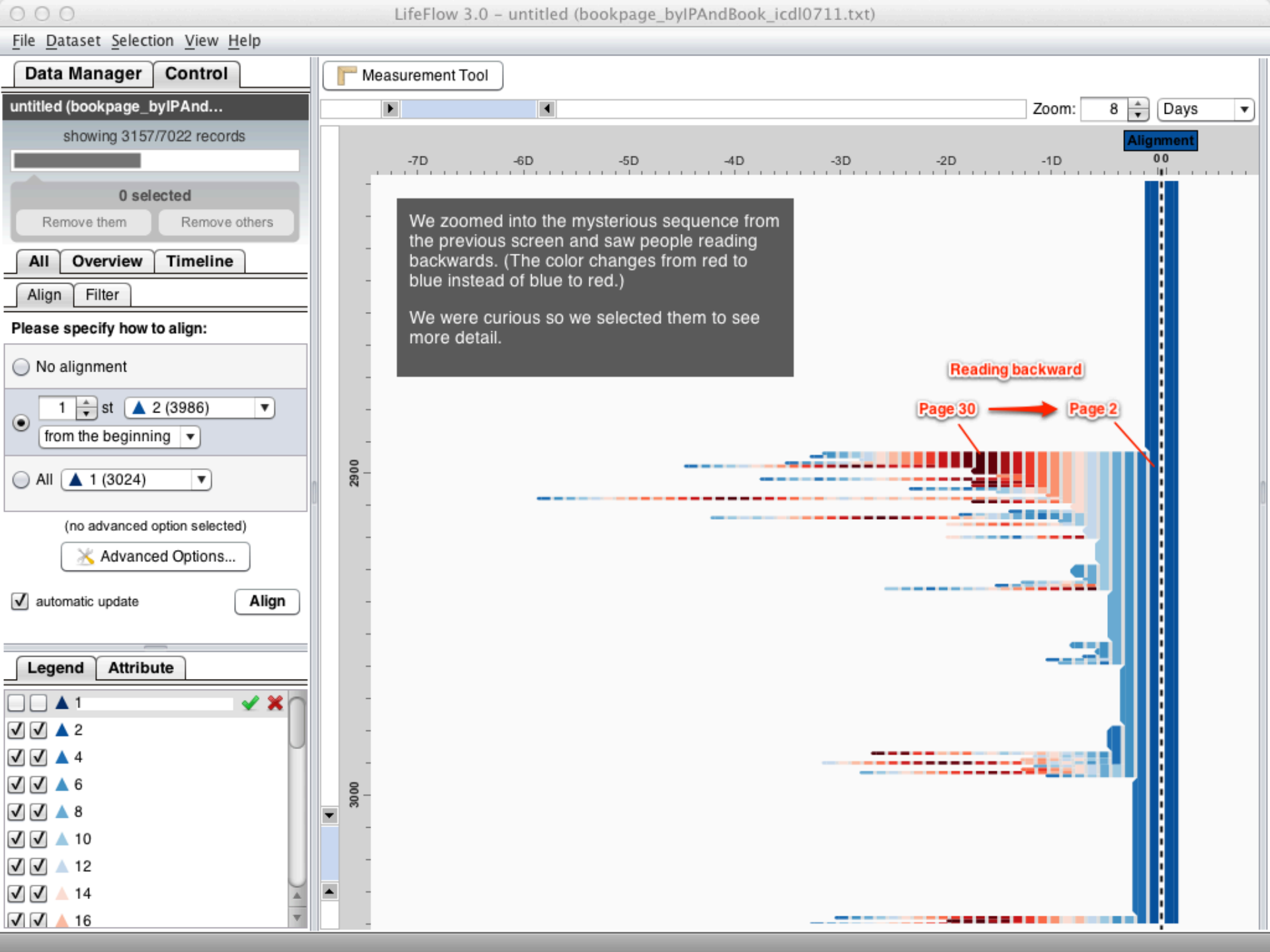
0
1000
2000
3000
4000
5000
6000

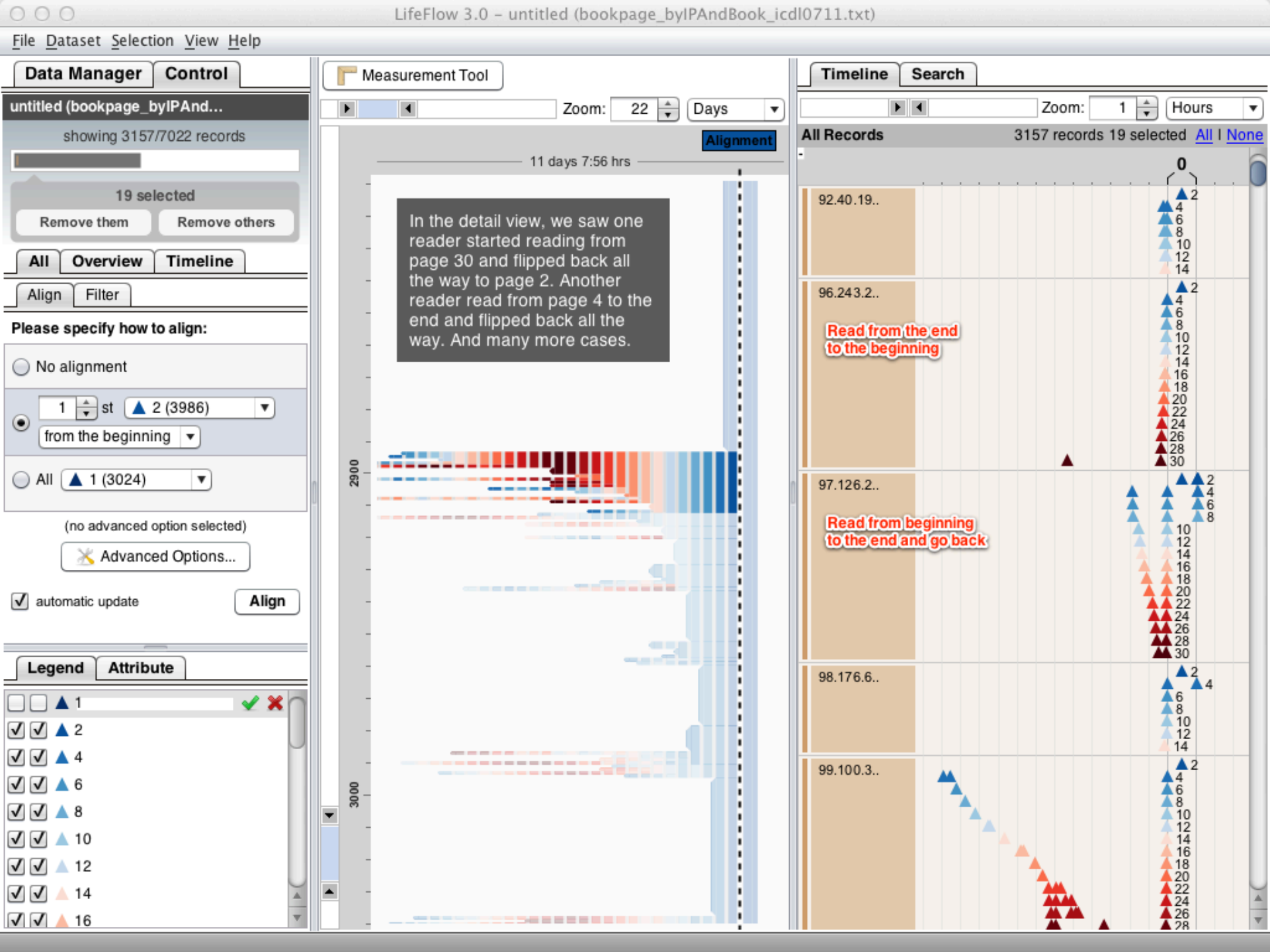
We encoded the page number using color gradient from blue to red.

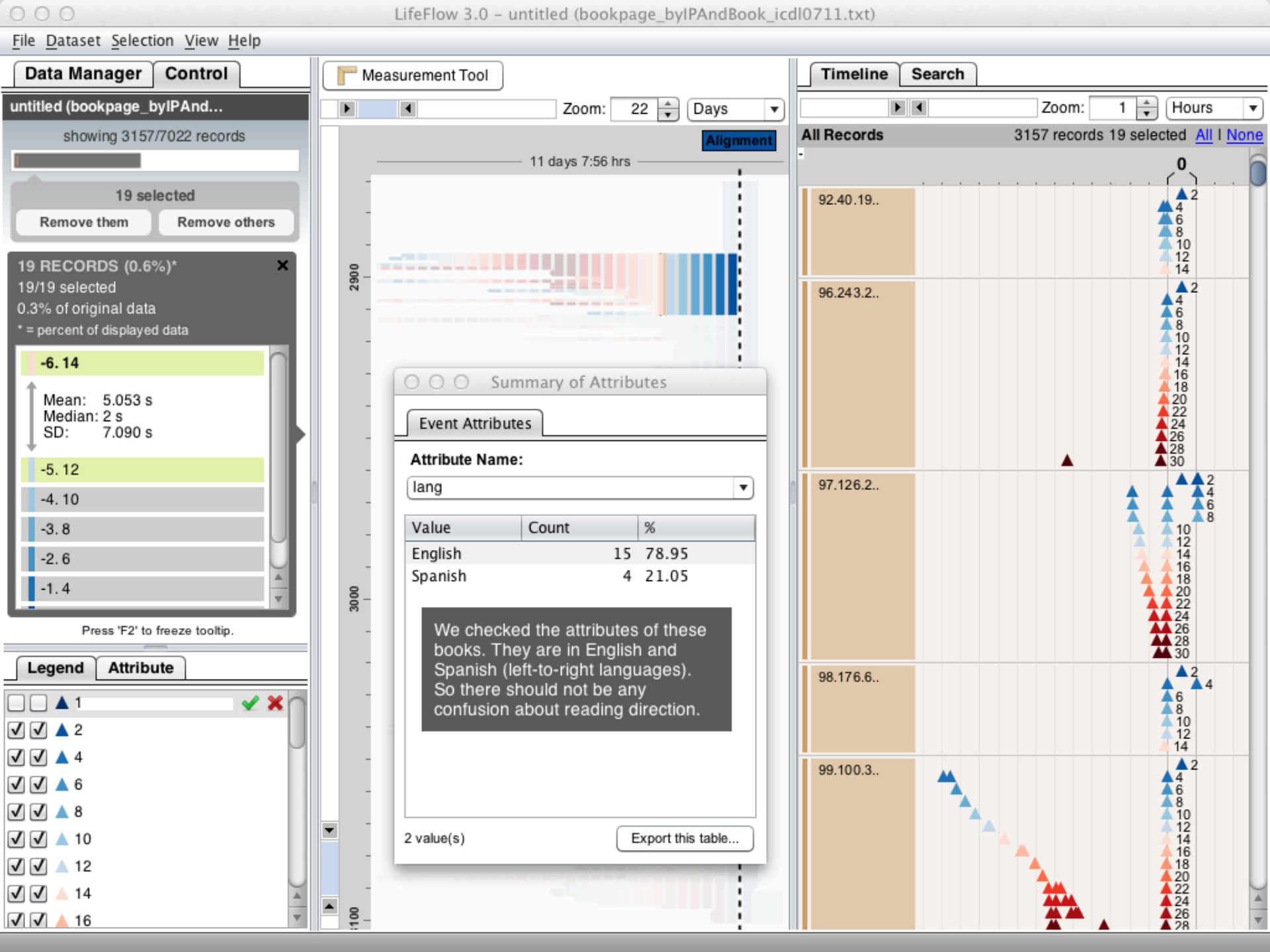
Because people also can the books in two mode:
single-page (1,2,3, ...)
and double-page (1,2,4,6, ...),
we hid the odd pages.

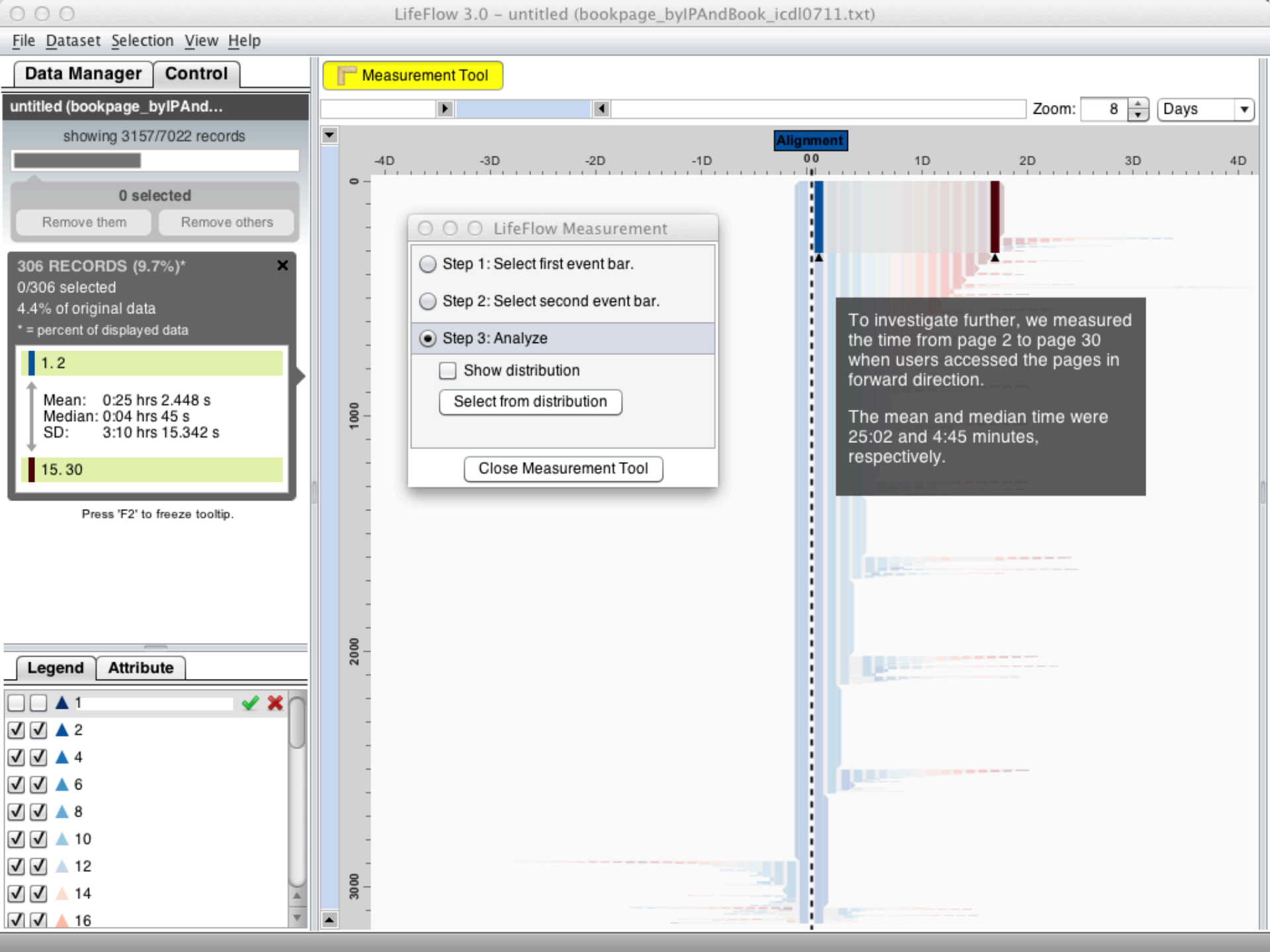

LifeFlow 3.0 - untitled (bookpage_byIPAndBook_icdl0711.txt)
File Dataset Selection View Help
Data Manager
Control
untitled (bookpage_byIPAnd...
showing 7022/7022 records
0 selected
Remove them
Remove others
All
Overview
Timeline
Readers started reading at different pages
Align
Filter
Please specify how to align:
No alignment
1
st
1 (3024)
from the beginning
All
1 (3024)
(no advanced option selected)
Advanced Options...
automatic update
Align
Legend
Attribute
1
2
4
6
8
10
12
14
16
Page #
Measurement Tool
Not everybody reads all pages, but mostly reads in order.
Zoom:
12
Days
Alignment
Page 1
Page 30
6 days 6:51 hrs
0
1000
2000
3000
4000
5000
6000
Page 2
Page 3
Most readers start reading from page 1
We found that people started their reading on different pages. Some started from the first page but some jumped into the later pages, probably skipping the empty pages in the beginning.
The height of the bars shows that people started on the earlier pages more than the later pages.
The color bars show a smooth gradient meaning that people mostly read in order. (Remember that the color coding is a gradient from blue to red.)

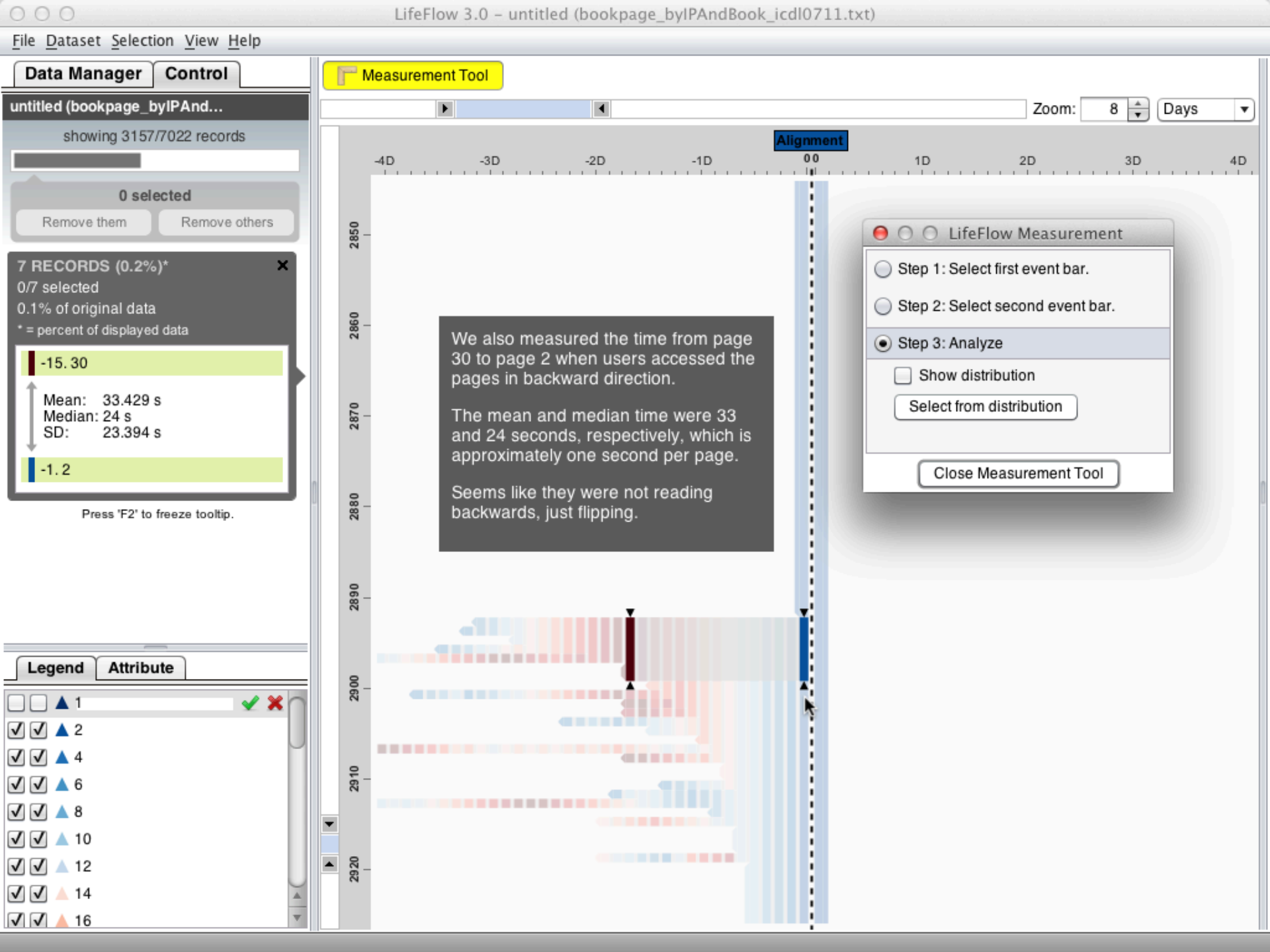

LifeFlow 3.0 - untitled (bookpage_byIPAndBook_icdl0711.txt)
File Dataset Selection View Help
Data Manager
Control
untitled (bookpage_byIPAnd...
showing 3157/7022 records
0 selected
Remove them
Remove others
All
Overview
Timeline
Align
Filter
Please specify how to align:
No alignment
1 st
2 (3986)
from the beginning
All
1 (3024)
(no advanced option selected)
Advanced Options...
automatic update
Align
Legend
Attribute
1
2
4
6
8
10
12
14
16
Measurement Tool
Zoom:
22
Days
Alignment
Aligned by page 2
Reading in order
We aligned by the second page to see how people read from the second page on.
The decreasing height of the bars as time went by shows that people gradually dropped out while reading.
The little long lines that split from the main trend in each step tell that some readers flipped back to the previous page before continued reading. The color changed to a little bit darker color (previous page) then changed back to the brighter colors (continue reading).
We also noticed some long patterns before the second page.
Flip back to previous page and continue reading
What is this pattern?
0
1000
2000
3000


LifeFlow 3.0 – untitled (bookpage_byIPAndBook_icdl0711.txt)
File Dataset Selection View Help
Data Manager
Control
untitled (bookpage_byIPAnd...
showing 3157/7022 records
0 selected
Remove them
Remove others
All
Overview
Timeline
Align
Filter
Please specify how to align:
No alignment
1 st
2 (3986)
from the beginning
All
1 (3024)
(no advanced option selected)
Advanced Options...
automatic update
Align
Legend
Attribute
1
2
4
6
8
10
12
14
16
Measurement Tool
Zoom:
8
Days
Alignment
-7D
-6D
-5D
-4D
-3D
-2D
-1D
00
2900
3000
We zoomed into the mysterious sequence from the previous screen and saw people reading backwards. (The color changes from red to blue instead of blue to red.)
We were curious so we selected them to see more detail.
Reading backward
Page 30
Page 2


LifeFlow 3.0 - untitled (bookpage_byIPAndBook_icdl0711.txt)
File Dataset Selection View Help
Data Manager
Control
untitled (bookpage_byIPAnd...
showing 3157/7022 records
19 selected
Remove them
Remove others
All
Overview
Timeline
Align
Filter
Please specify how to align:
No alignment
1 st
2 (3986)
from the beginning
All
1 (3024)
(no advanced option selected)
Advanced Options...
automatic update
Align
Legend
Attribute
1
2
4
6
8
10
12
14
16
Measurement Tool
Zoom: 22
Days
Alignment
11 days 7:56 hrs
In the detail view, we saw one reader started reading from page 30 and flipped back all the way to page 2. Another reader read from page 4 to the end and flipped back all the way. And many more cases.
2900
3000
Timeline
Search
Zoom: 1
Hours
All Records
3157 records 19 selected All | None
0
92.40.19..
96.243.2..
Read from the end to the beginning
97.126.2..
Read from beginning to the end and go back
98.176.6..
99.100.3..


LifeFlow 3.0 – untitled (bookpage_byIPAndBook_icdl0711.txt)
File Dataset Selection View Help
Data Manager
Control
untitled (bookpage_byIPAnd...
showing 3157/7022 records
19 selected
Remove them
Remove others
19 RECORDS (0.6%)*
19/19 selected
0.3% of original data
* = percent of displayed data
-6. 14
Mean: 5.053 s
Median: 2 s
SD: 7.090 s
-5. 12
-4. 10
-3. 8
-2. 6
-1. 4
Press 'F2' to freeze tooltip.
Legend
Attribute
1
2
4
6
8
10
12
14
16
Measurement Tool
Zoom: 22
Days
Alignment
11 days 7:56 hrs
2900
3000
Summary of Attributes
Event Attributes
Attribute Name:
lang
Value
Count
%
English 15 78.95
Spanish 4 21.05
We checked the attributes of these books. They are in English and Spanish (left-to-right languages). So there should not be any confusion about reading direction.
2 value(s)
Export this table...
Timeline
Search
Zoom: 1
Hours
All Records
3157 records 19 selected
All | None
0
92.40.19..
96.243.2..
97.126.2..
98.176.6..
99.100.3..


LifeFlow 3.0 – untitled (bookpage_byIPAndBook_icdl0711.txt)
File Dataset Selection View Help
Data Manager
Control
untitled (bookpage_byIPAnd...
showing 3157/7022 records
0 selected
Remove them
Remove others
306 RECORDS (9.7%)*
0/306 selected
4.4% of original data
* = percent of displayed data
1. 2
Mean: 0:25 hrs 2.448 s
Median: 0:04 hrs 45 s
SD: 3:10 hrs 15.342 s
15. 30
Press 'F2' to freeze tooltip.
Legend
Attribute
1
2
4
6
8
10
12
14
16
Measurement Tool
Zoom:
8
Days
Alignment
-4D
-3D
-2D
-1D
00
1D
2D
3D
4D
0
1000
2000
3000
LifeFlow Measurement
Step 1: Select first event bar.
Step 2: Select second event bar.
Step 3: Analyze
Show distribution
Select from distribution
Close Measurement Tool
To investigate further, we measured the time from page 2 to page 30 when users accessed the pages in forward direction.
The mean and median time were 25:02 and 4:45 minutes, respectively.


LifeFlow 3.0 – untitled (bookpage_byIPAndBook_icdl0711.txt)
File Dataset Selection View Help
Data Manager
Control
untitled (bookpage_byIPAnd...
showing 3157/7022 records
0 selected
Remove them
Remove others
7 RECORDS (0.2%)*
0/7 selected
0.1% of original data
* = percent of displayed data
-15. 30
Mean: 33.429 s
Median: 24 s
SD: 23.394 s
-1. 2
Press 'F2' to freeze tooltip.
Legend
Attribute
1
2
4
6
8
10
12
14
16
Measurement Tool
Zoom: 8
Days
Alignment
-4D
-3D
-2D
-1D
00
1D
2D
3D
4D
2850
2860
2870
2880
2890
2900
2910
2920
We also measured the time from page 30 to page 2 when users accessed the pages in backward direction.
The mean and median time were 33 and 24 seconds, respectively, which is approximately one second per page.
Seems like they were not reading backwards, just flipping.
LifeFlow Measurement
Step 1: Select first event bar.
Step 2: Select second event bar.
Step 3: Analyze
Show distribution
Select from distribution
Close Measurement Tool

Conclusion

We dug deeper into several records with backward access one-by-one and hypothesized a few possible scenarios.

- **Finished reading a part of the book**: For example, one reader read from page 10 to 18 which is exactly one chapter of the book, then flipped back to page 1 and did not have any interaction after that.
- **Shelf books**: ICDL has a membership system, which a member can store books on his/her virtual shelf for reading later. The system also remembers the last page read by the member. However, we assumed that some members might open the book and find it at page 30, so they tried to flip it back to the first page and start reading again from the beginning.
- **Parent preview**: Some users read a few pages then flipped back to the beginning and started reading from the beginning. We assumed that this is a behavior of parents selecting books for their children. They flipped a few pages to check that it is suitable then went back to the beginning to read it.

From these three possible scenarios, the users seem to depend heavily on the page-by-page (previous/next button) navigation. To change from page 30 back to page 1, they went to the previous page 30 times instead of using other controls that could let them jump to page 1 faster. This could be because the users could use their arrow keys to go to previous/next page, which was convenient for them, so they kept pressing the arrow keys, or the users did not know how to jump back to the first page directly.

However, to find the real explanation why many people access the online children books backwards or does the navigation need to be improved will require a further user study, which is beyond the scope of this case study. Therefore, we would like to take a step back and summarize what we learned from the ICDL web logs so far using LifeFlow.

LifeFlow showed potential for helping the ICDL administrator understand how people are accessing the online books. It highlighted the users' reading behaviors from the majority that read in order, the readers that skipped the the blank early pages right to the content, the readers that flipped back and continue reading, to the readers that flipped backwards all the way back to the first page. By understanding these behaviors better, the administrator can improve the website to suit the readers better.