Patterns and Sequences: Interactive Exploration of Clickstreams to Understand Common Visitor Paths

Supplemental Material: Designs from Earlier Iterations

Zhicheng Liu, Yang Wang, Mira Dontcheva, Matthew Hoffman, Seth Walker and Alan Wilson

Event Aggregation

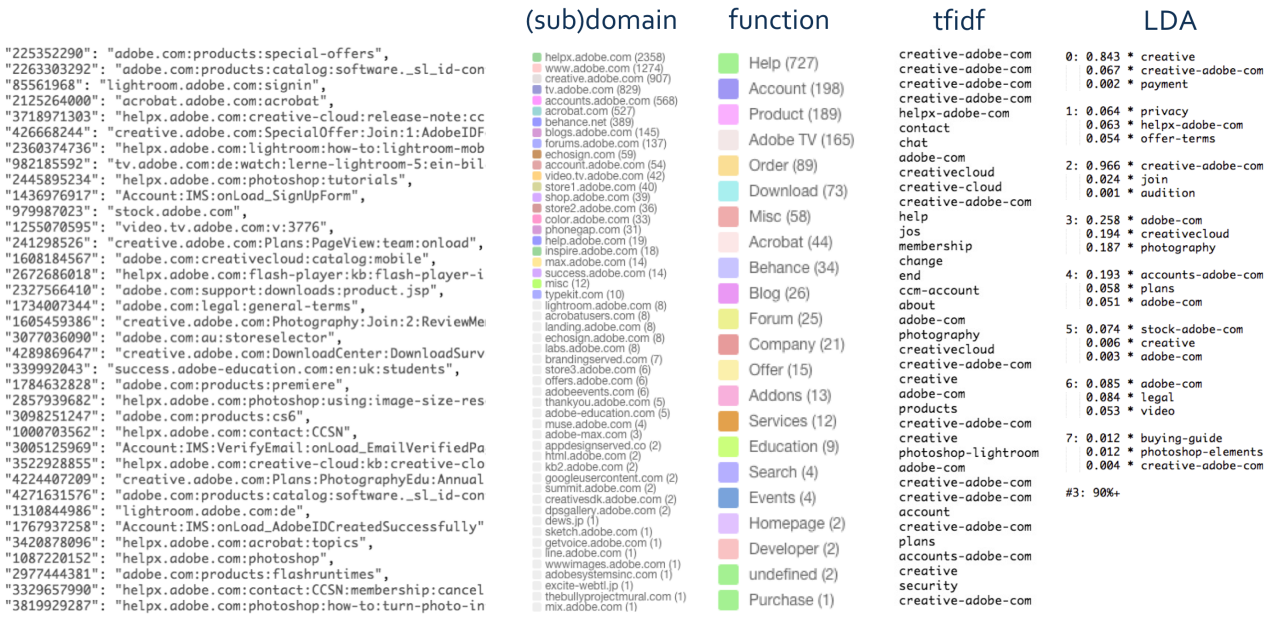

Figure 1: A demonstration of the event categorization results described in Section 5. From the left: an example set of event logs; event logs categorized and color coded by domain names (URL hierarchies); event logs categorized and color (manually) coded by functions based on analysts domain knowledge; top term frequency results based on the tfidf model; an example of the eight "top" clusters generated by the LDA model; Due to the characteristics of the text information encoded in event logs, automatic event categorization did not work well in our scenario.

Sequence Clustering

Figure 2: A screenshot of an early prototype. Using the side panel on the left, users can assign/remove color to event categories. On the right, the event sequences are visualized as vertically stacked squares with colors representing the categories. The event sequences are grouped to help understand to provide visual summaries of different clustering results. Although at a glance the users were able to obtain a high-level understanding of the characteristics of each cluster, no further insight could be derived from this view.

Sequence Clustering: Vertical Stacking

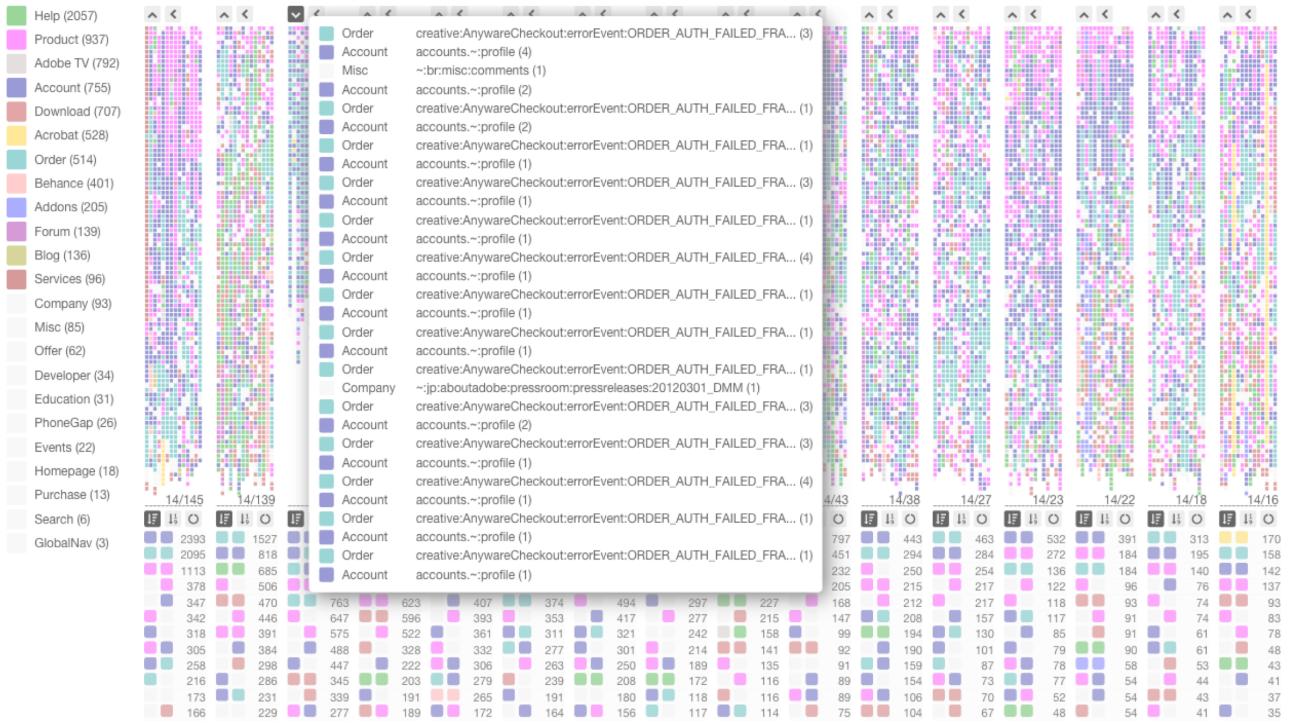

Figure 3: A demonstration of the rich interactions of the above prototype. Besides assigning colors, users can hover and reveal detail event sequence information; the sequence can be stacked both vertical and horizontally, by merging events of the same category. we also reveal the most frequent transitions between pages in the bottom panel to facilitate the event summarization.

Motif Extraction

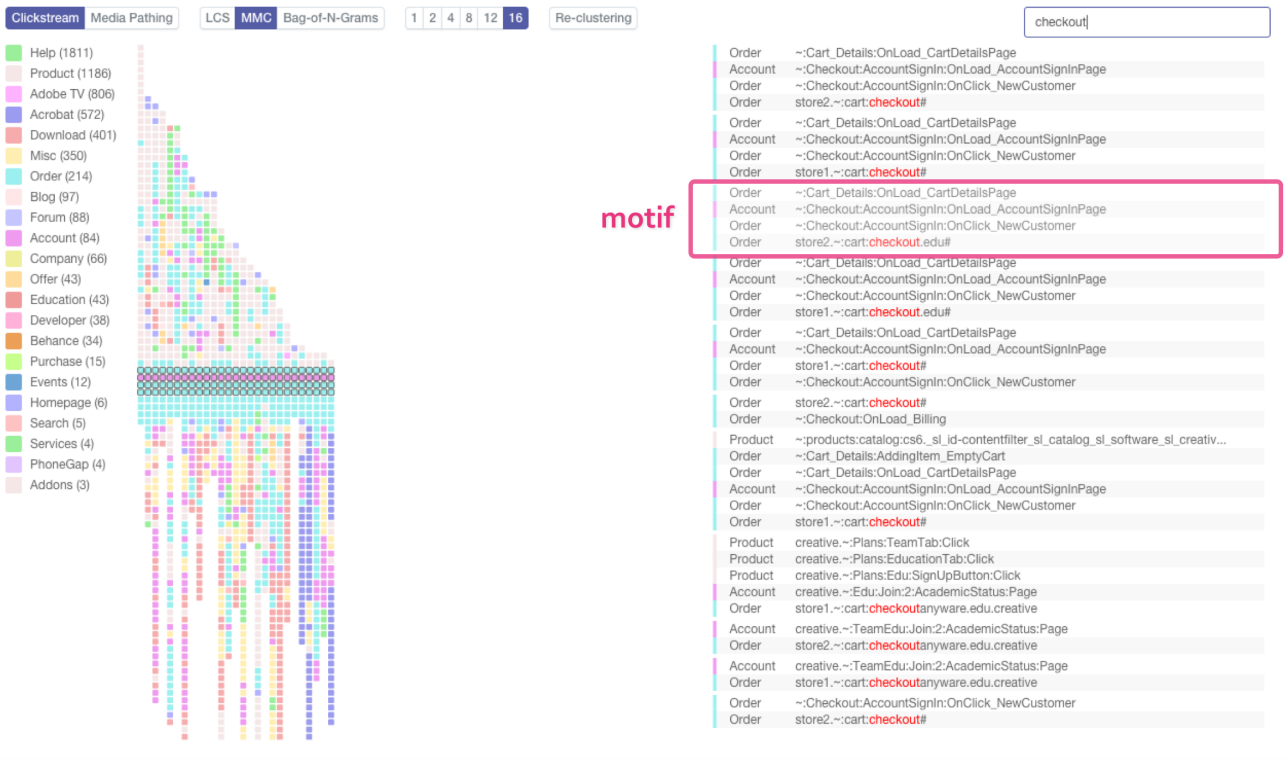

Figure 4: The second design iteration based on the idea that key motifs in the funnel can help summarizing the event sequences. As shown in the figure, a filter panel was added to the right to surface frequent motifs per users' input. The event sequences are then filtered and aligned to the selected motif, leading to a cleaner view of what is going on before and after the motif.

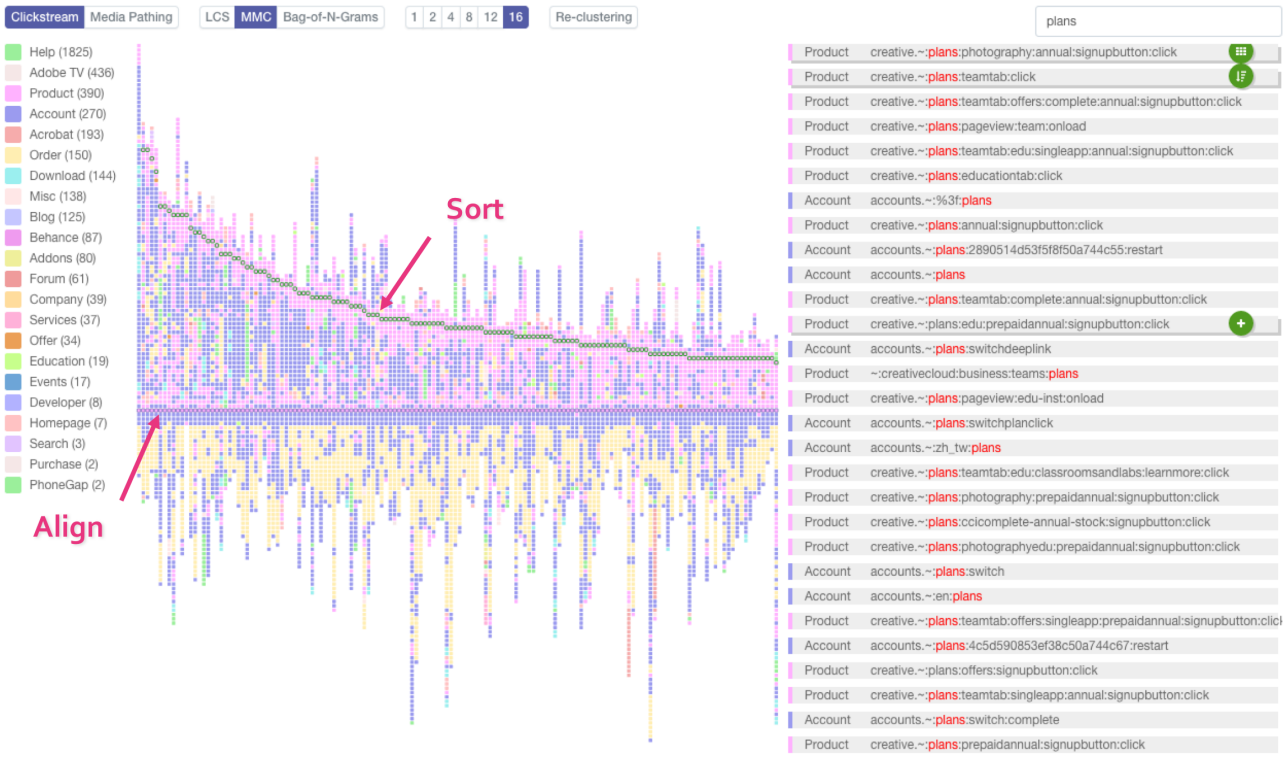

Figure 5: In addition to single motif alignment, sorting by a second motif helped revealing further information. As shown in the figure, a clear pattern of large portion of yellow regions (type ORDER events) started to show up below the purple band (Signup events), and the portion of events between the green line (first production page view) and the purple band identifies users' behaviors between this two key motifs within the conversion funnel.

Segmenting Sequences by Motifs

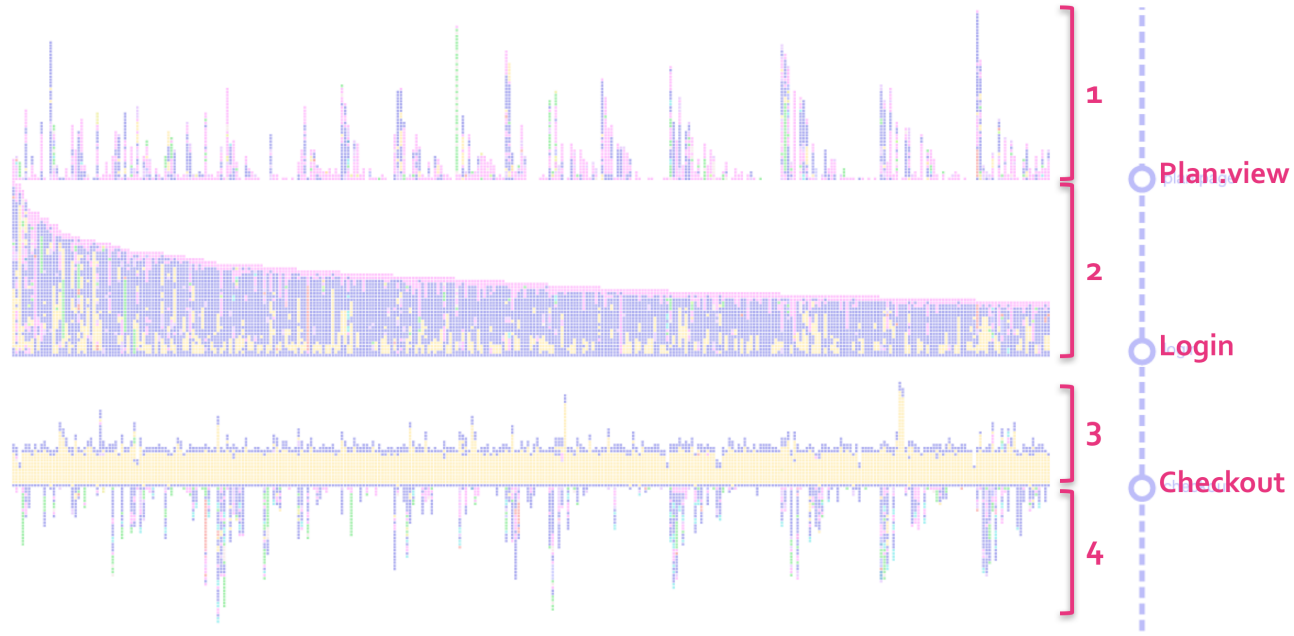

Figure 6: A demonstration of event sequence segmentation based on multiple key motifs. From top to bottom, 1: events after users open the company website and before they visit the product page; 2: user trails after visiting the production page and before the login page, which was considered most important to the analysts; 3: the highly analogous checkout process; 4: events after user's purchase.

Sequential Patterns

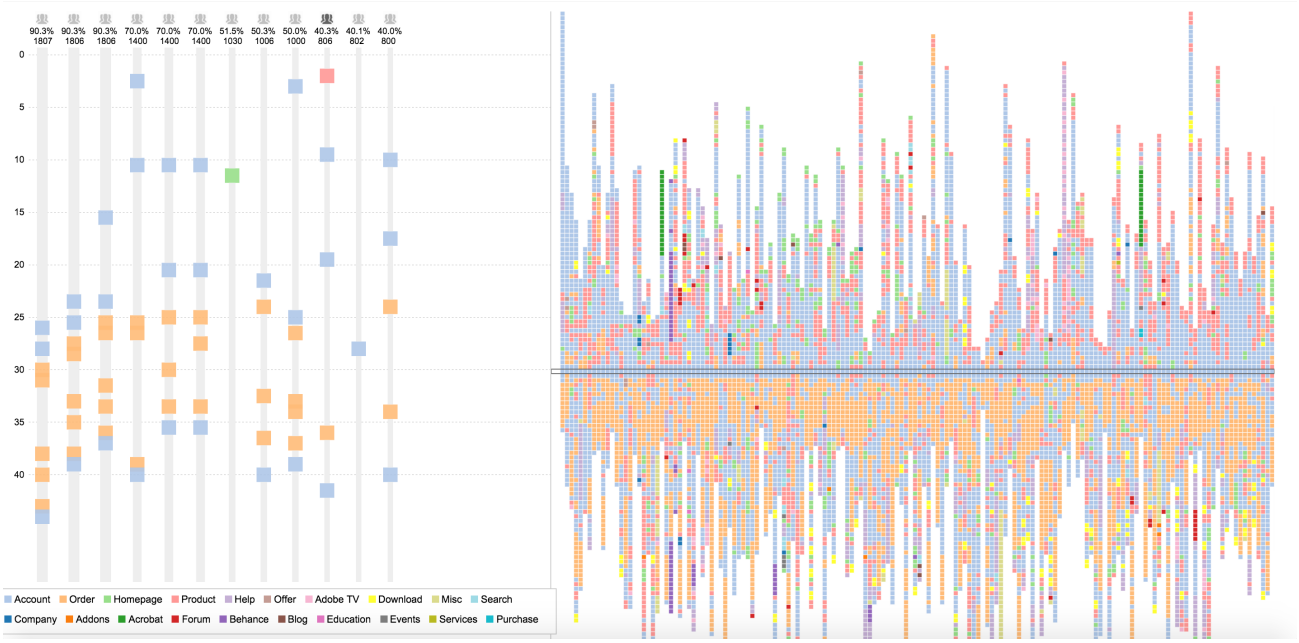

Figure 7: An early design for the pattern view and the sequence view. Here we try to fit all the patterns in a single overview. Analysts commented that it was difficult to distinguish the patterns visually and wanted to see the event labels by default. We thus revised the design to space out the patterns and display event names (Section 9.1).

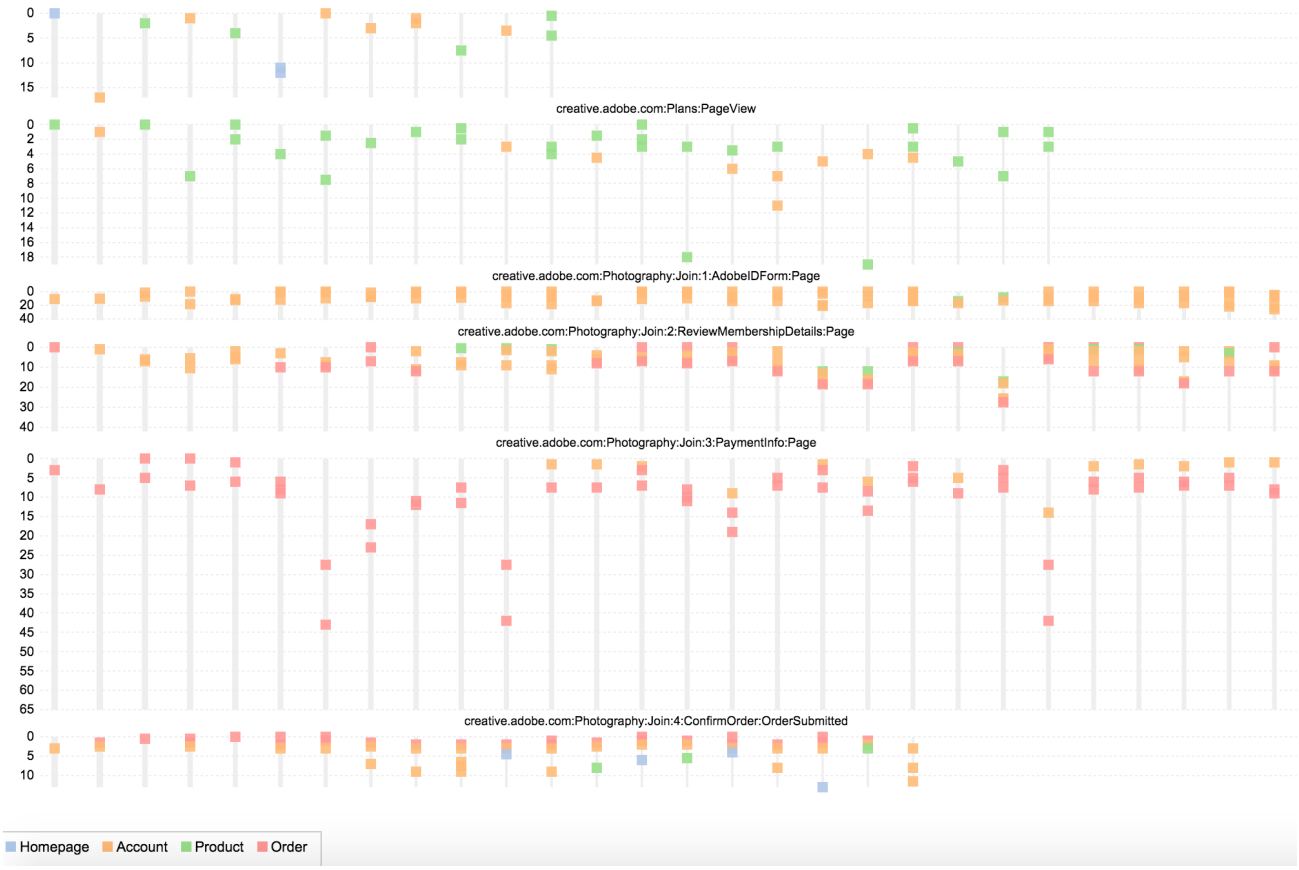

Figure 8: An early design where the sequences are broken down by key events specified by the analysts, and the pattern view shows patterns for the subsequences in each segment.