JsonCurer: Appendix

This appendix discusses additional details related to the visual design of JsonCurer, including summary visualizations and a node reordering algorithm, as mentioned in Section 6.1.

1. Summary Visualization

We utilize summary visualizations tailored to each data type to help users gain a deeper understanding of their data and effectively identify potential quality issues. These visualizations are presented in Table S1.

Table S1. The summary techniques and charts for different data types. The size of an *array*/*dict* is its hierarchy depth and the number of the elements of the *array*/the k-v pairs of the *dict*.

Data Type	Summary Technique	Chart Type
Str	Category frequency	Bar Chart
	String length frequency	Histogram
Num	None	Box Plot
	Bin frequency	Histogram
Bool	True/False frequency	Pie Chart
Null	Count	Bar Chart
Array	Array size frequency	Heat Map
Dict	Dict size frequency	Heat Map

2. Node Reordering Algorithm

We notice that nodes within some bubbles may be spaced apart (see Figure S1A), leading to the intersection of bubbles and separation of focuses. To alleviate this problem, we employ a node reordering algorithm that aims to minimize the vertical distance between the nodes of each bubble (see Figure S1B). This algorithm focuses on two specific aspects, namely minimizing the vertical distance between each two brother nodes in the brother issue sets and minimizing the vertical distance between each child node and the center point in the parent-child issue sets. In our algorithm, a brother issue refers to an issue involving only sibling nodes, while a parent-child issue involves both parent and child nodes. For example, there are three brother issue sets (i.e., {C1, C3, C4}, {C5}, and {C3, C6}) and one parent-child issue set (i.e., {P0, C6}) in Figure S1A. The center point of the child nodes in the {P0, C6} issue set is calculated as $(1+6)/2 = 3.5$. Mathematically, we formulate the minimization problem as follows:

$$\min\left(\sum_{s}^{S_b}\sum_{i,j}^{s}|i-j| + \sum_{s_c}^{S_{pc}}\sum_{i}^{s_c}|i - Center_p|\right) \tag{S1}$$

Here S_b is the collection of brother issue sets, and s is the vertical point in the arrangement order of each brother node (e.g., the vertical point of C1 is 1). S_{pc} is the collection of parent-child issue sets, and s_c is the vertical point of each child node (e.g., the vertical point of C6 is 6). $Center_p$ is the center point of the child nodes (e.g., 3.5). We utilize a greedy strategy that exchanges two node positions to optimize this minimization goal.

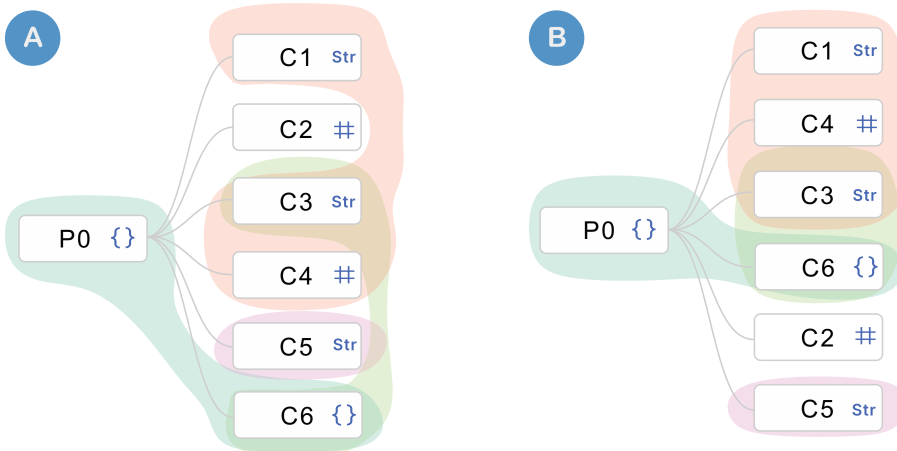

Fig. S1. An example of reordering nodes from A to B based on our algorithm. The vertical distance between the nodes of bubbles is decreased from 11.5 (A) to 5.5 (B). The detailed calculations are presented in Table S2.

Table S2. The vertical distance between the nodes of bubbles.

Tree	Brother Issue Sets			Parent-Child Issue Sets	Vertical Distance
	{C1, C3, C4}	{C5}	{C3, C6}	{P0, C6}	
A	$(3-1)+(4-1)+(4-3)=6$	0	$6-3=3$	$6-3.5=2.5$	11.5
B	$(2-1)+(3-1)+(3-2)=4$	0	$4-3=1$	$4-3.5=0.5$	5.5