We designed and conducted two user study evaluating our tree parsing and clustering visualization tool as follow.

Participants: We asked 10 graduate student in computer science and engineering department to participant in our user study. They are decided into two groups, with 5 English native speakers and 5 using English as second language.

Instructions: each participant is given an example on each tree parsing and clustering. And given 5 minute to learn the example and ask whatever question they have.

Experiment procedure:

1. Clustering: Each of them are given 10 articles for clustering with 5 conducted using our visualization tool (figure 5-1)and 5 conducted using Excel in the traditional way, where they will be clustering expressions by putting them into the same row (fiture 5-2).The order of the article and the tool they use will be randomly ordered.
2. Clustering: Each of them are given 10 sentences for tree parsing with 5 conducted using our visualization tool (Figure 5-3)and 5 conducted using plain text editor using bracket in the traditional way, where they will be doing sentence parsing by using bracket to separate different elements in an sentence. (Figure 5-4).The order of the article and the tool they use will be randomly ordered.

Data collection & Evaluation:

Both time consumption and task accuracy will be evaluated. The accuracy is defined as follow:

1. Clustering accuracy: connect all the expressing in each clustering. And count all the connections in the graph. Calculate the correct connection in the clustering result and clustering accuracy= correct connections/all connections
2. Clustering recall
3. Tree parsing accuracy
4. Tree parsing recall
5. Clustering time: time consumption per word is calculated
6. Tree parsing: parsing time per word is calculated

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Participant | 1# | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Parsing time/word |  |  |  |  |  |  |  |  |  |  |
| Recall |  |  |  |  |  |  |  |  |  |  |
| Acuracy |  |  |  |  |  |  |  |  |  |  |
| Clustering time/word |  |  |  |  |  |  |  |  |  |  |
| Recall |  |  |  |  |  |  |  |  |  |  |
| Accuracy |  |  |  |  |  |  |  |  |  |  |

Evaluation: