We designed and conducted two user studies evaluating our tree parsing and clustering visualization tool as follow:

Participants: We asked 6 graduate students aged 21 to 26 years old to participant in our user study. 3 are native English speakers, the other 3 learned English as a second language. We divide them into native group and foreign group.

Instructions: each participant is given two answer examples one on tree parsing and one on 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.

Evaluation:

Clustering:

Methods: time efficiency and result quality will be evaluated.

Time efficiency is defined as time user per word.

Results quality is evaluated on two aspects:

1. Purity defined as et Nij be the number of objects in cluster i that belong to class j, and let Ni = Cj=1 Nij be the total number of objects in cluster i. Define pij = Nij/Ni; this is the empirical distribution over class labels for cluster i. We define the **purity** of a cluster as pi 􏰪 maxj pij, and the overall purity of a clustering as purity 􏰪 􏰀 Ni pi iN
2. Let U = {u1,...,uR} and V = {v1,...,VC} be two different partitions of the N data points, i.e., two different (flat) clusterings. For example, U might be the estimated clustering and V is reference clustering derived from the class labels. Now define a 2 × 2 contingency table containing the following numbers: TP is the number of pairs that are in the same cluster in both U and V (true positives); TN is the number of pairs that are in the different clusters in both U and V (true negatives); FN is the number of pairs that are in the different clusters in U but the same cluster in V (false negatives); and FP is the number of pairs that are in the same cluster in U but different clusters in V (false positives). A common summary statistic is the **Rand index**:

R TP+TN TP +FP +FN +TN

Calculated results of the average of the stated variable are shown as below：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Participant | 1# | 2# | 3# | 4# | 5# | 6# |
| Text Editor | time |  |  |  |  |  |  |
| Purity |  |  |  |  |  |  |
| Rand index |  |  |  |  |  |  |
| Clustering Tool | Time efficiency |  |  |  |  |  |  |
| Purity |  |  |  |  |  |  |
| Rand index |  |  |  |  |  |  |