

Appendix I: R results and Example Usage

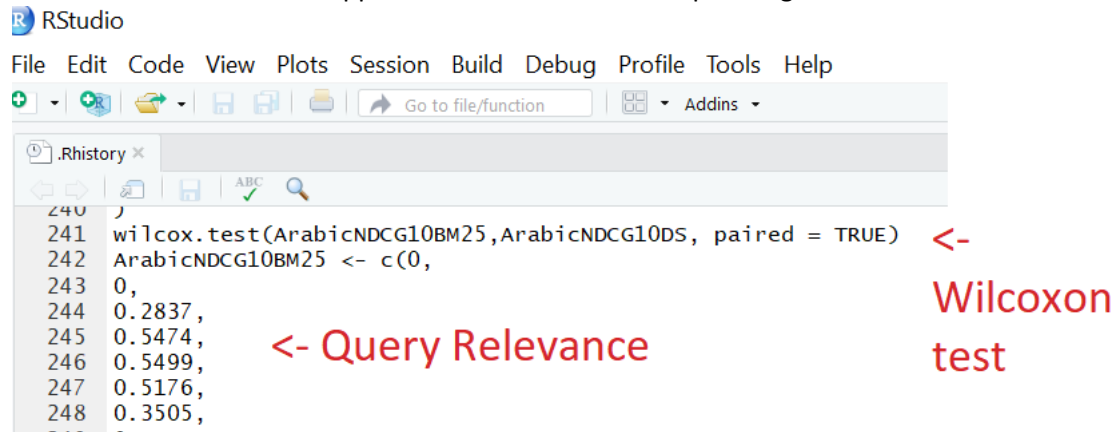


Figure 1: R querying with Wilcoxon test

```
[Workspace loaded from ~/.RData]  
  
> wilcox.test(ArabicNDCG10BM25,ArabicNDCG10DS, paired = TRUE)  
  
    Wilcoxon signed rank test with continuity correction  
  
data:  ArabicNDCG10BM25 and ArabicNDCG10DS  
V = 158, p-value = 0.003595  
alternative hypothesis: true location shift is not equal to 0
```

Figure 2: P-value results

Synopsis of what is happening above:

Figure 1 depicts RStudio, the software that was used to do the statistical significance testing. As can be seen, Arabic for metric NDCG and retrieval algorithm BM25 is used against metric NDCG and retrieval algorithm Dirichlet Smoothing (DS). The P value in figure 2 is the significance, in this case 0.003595, which is significant as it is less than 0.05.