

# Errata for

## “Querying Text Databases for Efficient Information Extraction”

as published in the Proceedings of the IEEE ICDE 2003 Conference

Due to an error in the implementation of the *Tuples* strategy described in [AG03], not all of the tuples were used for querying for the *DiseaseOutbreaks* relation. (The implementation of *Tuples* for the *Headquarters* relation was correct.) Therefore, the recall and precision of the *Tuples* strategy were incorrectly reported in Figure 12 (a) and (b) of [AG03]. The correct value for recall of *Tuples* for all the values of *MaxFractionRetrieved* that we reported is 46.1%, and the correct precision is 28.9%. The correct version of Figure 12 of [AG03] is shown below.

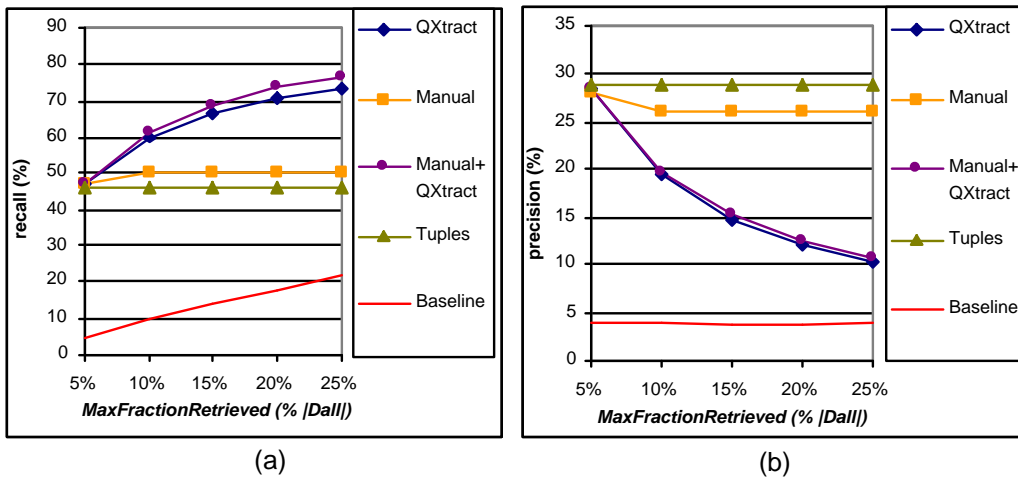


Figure 1: Recall (a) and precision (b) of *QXtract*, *Tuples*, *Manual*, *Manual+QXtract*, and *Baseline* over the **test** database using *Proteus* as the target information extraction system (*DiseaseOutbreaks* relation).

Figure 1 shows that the general conclusions reported in [AG03] on the comparison of *Tuples* against the other strategies still hold, but *Tuples* exhibits higher recall than previously reported, while precision is correspondingly lower.

## References

[AG03] Eugene Agichtein and Luis Gravano. Querying text databases for efficient information extraction. In *Proceedings of the 19th IEEE International Conference on Data Engineering (ICDE)*, 2003.