**Extraction Scoring Diff System**

**1. Inputs**

1. 2 score reports to diff; eg. 3-extraction-scoring and 4-extraction-scoring

**2. Outputs**

1. a diff report; eg. 3-4-extraction-scoring-diff

**3. Arguments**

1. (optional) args(0) can be an integer which is used to determine the input files and the output file; eg., args(0) = 4 would read in 3-extraction-scoring and 4-extraction scoring, and write 3-4- extraction-scoring-diff.

2. (optional) args(0) and args(1) can be integers which determine the input and output file names;

eg. 1: args(0) = 4 and args(1) = 6 would input

4-extraction-scoring and 6-extraction-scoring

and would output

4-6-extraction-scoring-diff.

eg. 2: args(0) = 6 and args(1) = 4 would input

6-extraction-scoring and 4-extraction-scoring

and would output

6-4-extraction-scoring-diff.

**4. How to Run**

0. edit ImplicitRelationExtractor/src/main/resources/extraction-scoring-diff.conf

to point to your desired input and output files

1. cd ImplicitRelationExtractor

2. sbt

3. sbt> runMain testers.ExtractionScoringDiff

(creates (n-1)-n-extraction-scoring-diff, where n is determined by the number in the file, file\_sequence\_number\_diff specified in extraction-scoring-diff.conf)

or

4. sbt> runMain testers.ExtractionScoringDiff 6

(creates 5-6-extraction-scoring-diff)

or

5. sbt> runMain testers.ExtractionScoringDiff 5 6

(creates 5-6-extraction-scoring-diff)

**5. Assumptions**

1. Input Files:

1. Ignored Lines:

-input lines which don't start with an integer are ignored as extractions;

e.g.,a header line of column names can be included and will be ignored

2. Sequence Number:

-file\_sequence\_number\_diff = n, is used to read in:

(n-1)-extraction-scoring and

n-extractionscoring,

and to write:

(n-1)-n-extraction-scoring-diff

3. To pick-up the Summary Stats Tables, it assumes that the table has 4 columns ` separated by double-tabs (“/t/t”). The table starts when the first token of a line is

equal to “Correct” and ends when the last (4th) token of a line equals “total”.

**6. Report Content**

1) the names of the two scoring reports

2) the correct extractions in the first report not in the second

3) the correct extractions in the second report not in the first

4) the incorrect extractions in the first report not in the second

5) the incorrect extractions in the second report not in the first

6) the summary stats table from the first report

7) the summary stats table from the second report