

First Experiences with TIRA for Reproducible Evaluation in Information Retrieval

Tim Gollub, Steven Burrows, Benno Stein

PAN

Competition on plagiarism detection hosted at CLEF.

- ▶ Task definition: Given a pair of suspicious and source document, record all passages in the suspicious document that are plagiarized from the source document.

Suspicious Document d_{plg}

Source Document d_{src}

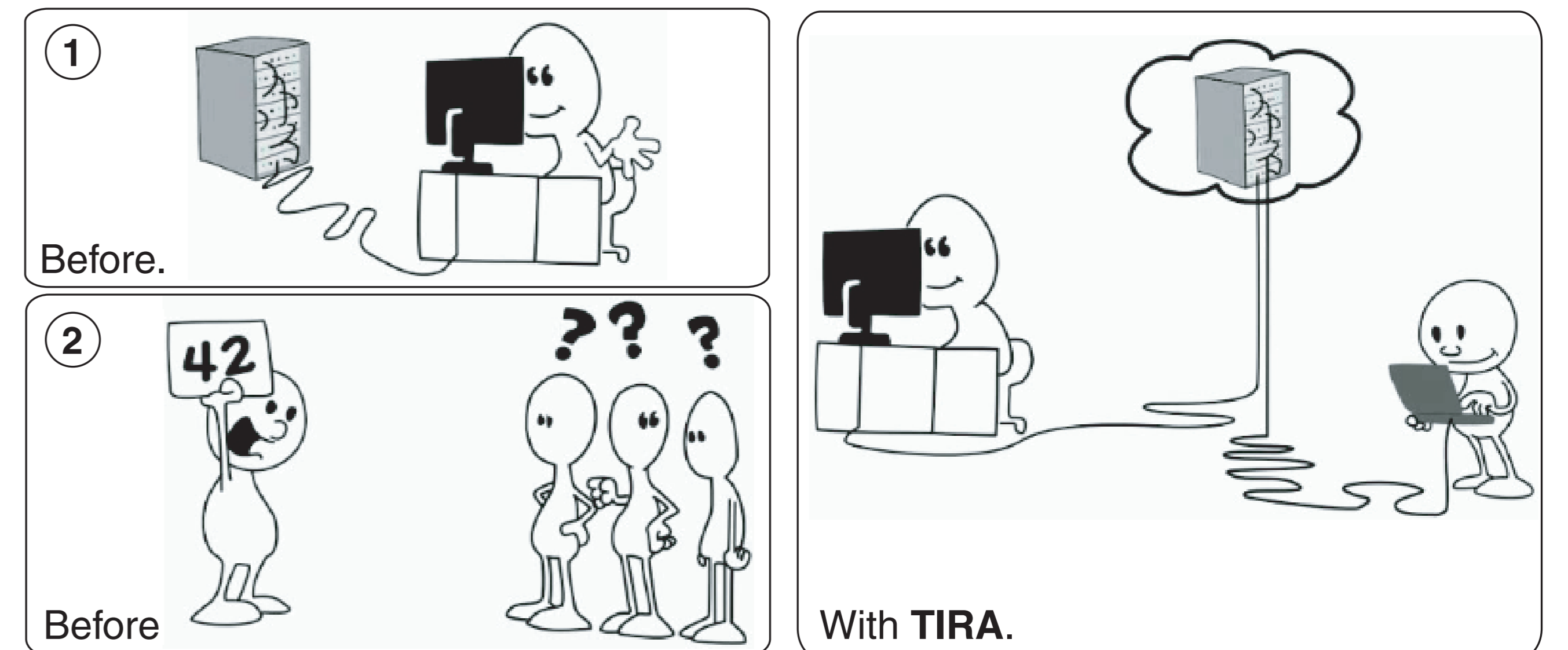
- ▶ Evaluation metric is the *plagdet* score:

$$plagdet(Det, Truth) = \frac{F_1(Det, Truth)}{\log_2(1 + granularity(Det, Truth))}$$

TIRA

Open source experimentation platform for IR research.

- ▶ Runs experiments and stores results.
- ▶ Disseminates experiments and results as a web service.
- ▶ Instances can be connected to a TIRA network.
- ▶ Supports organization of competitions.
- ▶ Allows public experiments on private data.



TIRA @ PAN2012

1. Training Service

pan2012/training-service

- Detection Upload File +
- Team +
- Testset +

Search Execute

```
tira@node1:~$
unzip -o $Detection -d det &&
python $PROGRAM/perfmeasure.py
-p /pan12-training-sets/$Testset/
-d det > scores.txt
```

| Detection | Team | Testset | Precision | Recall | Gran | Plagdet | Avg Time | Documents | Output Dir |
|------------------|-------|-------------------------|-----------|--------|--------|---------|----------|-----------|------------|
| detection-42.zip | webis | 06_simulated_paraphrase | 0.8881 | 0.5363 | 1.0014 | 0.6680 | 6.2120 | 500 | output-dir |

- ▶ Participants upload detection results for a given training set.
- ▶ From the user inputs the program execution command is generated through substitution.
- ▶ Detection results are unzipped and evaluated with an implementation of *plagdet*.
- ▶ Participants receive performance results in a result table.
- ▶ The training service served as a leaderboard during the competition.

2. Evaluation Service

pan2012/evaluation-service

- Testset +
- Detector +
- Detector -
- Detector -

Search Execute

Windows 7

Ubuntu 12.04

| Testset | Detector | Precision | Recall | Gran | Plagdet | Avg Time | Documents | Output Dir |
|---------|--------------|-----------|--------|--------|---------|----------|-----------|------------|
| 08_all | kongleilei12 | 0.8249 | 0.6782 | 1.0109 | 0.7386 | 5.9187 | 3033 | output-dir |
| 08_all | kasprzak12 | 0.8931 | 0.5524 | 1.0000 | 0.6826 | 5.3679 | 3033 | output-dir |
| 08_all | torreon12 | 0.8344 | 0.5004 | 1.0009 | 0.6252 | 0.1900 | 3033 | output-dir |

- ▶ TIRA platforms are provided for two operating systems.
- ▶ Participants submit their plagiarism detection software for deployment on the appropriate instance.
- ▶ A third TIRA instance evaluates all submissions on the private test set and provides the overall results.

TIRA System Architecture

