

metahtml: a library for extracting structured information from webpages

Anonymous EMNLP submission

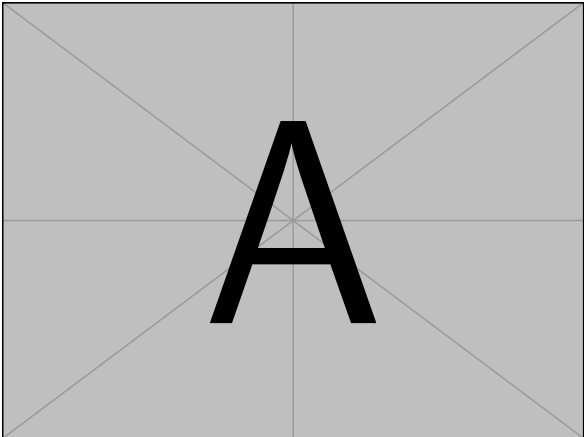


Figure 1: This is the summary figure of our paper.

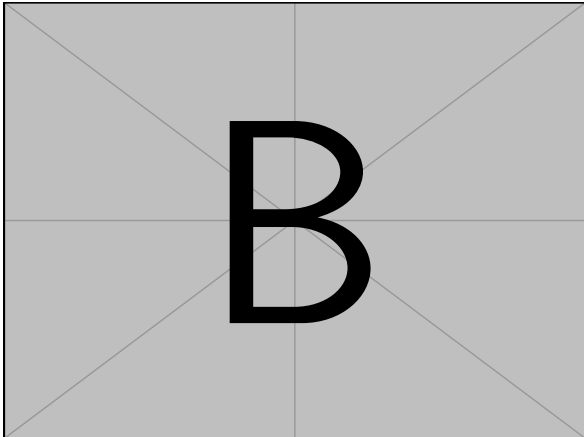


Table 1: Summary of our collected dataset.

Abstract

**FIXME:** the abstract should be written last

1 Introduction

**FIXME:** the intro should be written last, but you should start thinking about which citations should get mentioned here

1.1 Contributions

**FIXME:** add the contributions; this should be the first thing to focus on, since exactly what our contributions are will determine how the rest of the paper gets written; but we will redo the exact wording of this when we write the intro properly

1.

2.

2 Related Work

**FIXME:** Find 5-20 papers that you think are related to this research. For each paper, include the citation and 1-5 sentences explaining why that paper is important. At this point, these summaries

should just be in outline format. Later, we will go through and put this together into a nice narrative.

Notice the correct way to use the cite and citet commands in the following examples:

NLP is awesome (Aho and Ullman, 1972).

Aho and Ullman (1972) show that NLP is awesome.

3 The Metahtml Algorithm

**FIXME:** probably you're not familiar enough with the algorithm yet to write this section, so focus on the data collection part first

4 Data Collection

**FIXME:** this is the main section that can be written right now; do your best to write as much of this as you can; once you're done, we'll go over what you've written; probably most of it will have to be rewritten, but that's 100% normal even for experience writers

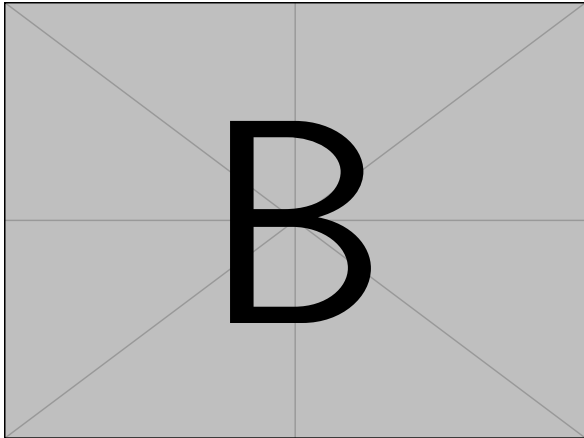


Table 2: The runtimes of the different libraries.

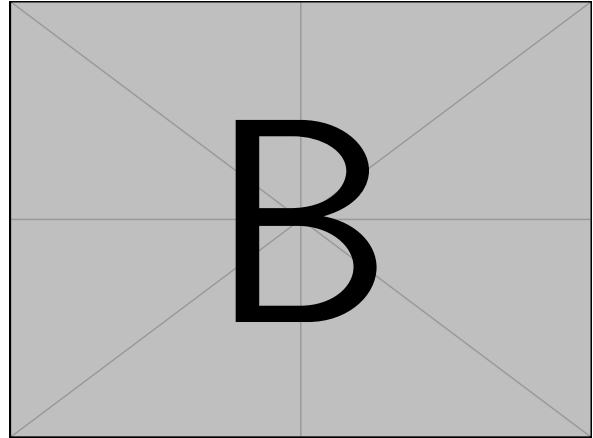


Table 5: The runtimes of the different libraries.

annotators	surachanliao	Tonnpo	joeybodoia	mikeizbicki
surachanliao	100.00	87.18	90.30	90.23
Tonnpo	88.80	100.00	93.62	93.76
joeybodoia	90.30	91.92	100.00	94.82
mikeizbicki	90.23	92.06	94.82	100.00

Table 3: Inner-annotator agreement of the labeled text dataset.

#### 4.1 Dates

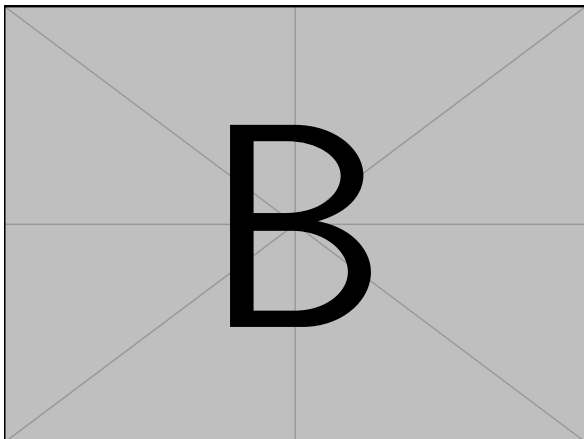
#### 4.2 Text

### 5 Results

### 6 Conclusion

### References

Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.



**FIXME:** this table will be automatically loaded from the file fig/results.txt; you should adjust your benchmark script so that it can automatically generate this file

Table 4: The results of applying metahtml to our manually annotated dataset.