Review of Pronominal Anaphora Resolution in NLP by Johan Winge

The paper starts with a long introduction to the main keywords of the paper; coreference and anaphora. This is good, because it explains the concepts of the terms thoroughly, along with helpful examples. Later on in the paper, two similar systems are chosen for comparisons. A reason for the choice would have been informative, especially since the two systems are so similar.\(^1\) What more, the paper could have benefited from adding a more different system, for example one trained on a corpus to provide a contrast against the two current systems.

The testing data for the evaluation in the paper is a small corpus consisting of six short texts of different genres. While the approach of finding out what kinds of mistakes the systems make, rather than the overall results, is acceptable, some kind of evaluative measure (precision, recall, F1-score) would have been of good use, since the actual comparison is lacking. After reading the paper, the, according to me, important question of a comparative evaluation ("how did they perform in comparison to each other?") remained unanswered. Of course, any general conclusions are better to leave for other, bigger studies, but even comparative results of such a small study is better than no results at all.

Overall when it comes to the contents of the paper, I have few things to complain about. Especially the introduction part was, as mentioned, very well written and informative. The testing part was lacking a more different system, and the results was lacking an actual comparative evaluation.

When it comes to the form of the paper, I feel that the text would benefit from being fully justified (like this text is), instead of simply left-aligned. This helps the paper looking a lot more professional. Secondly, the paper contains a lot of long sentences, many of them being 4 or more rows long. While this is not always a problem, it arguably makes the sections containing more difficult and technical information somewhat more difficult to grasp. Lastly, I could find a handful of spelling mistakes. While a majority of them was simple errors that some more thorough spell checking could solve, there was one case of mixed tenses in the abstract, which might be a little more difficult to spot.

\(^1\) From Winge's presentation for the paper, I remember mentioning that the two systems were pretty much chosen at random. But still, that part should have been in the paper as well.