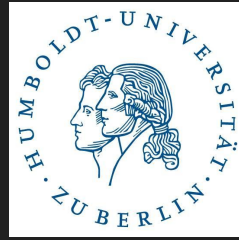


{headtcentre}

humboldt-elsevier
advanced data & text centre



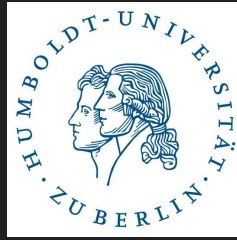
Forschung zu Wissenschaftlichem Fehlverhalten am HEADT Centre

Michael Seadle

Vision Statement

{headtcentre}

humboldt-elsevier
advanced data & text centre



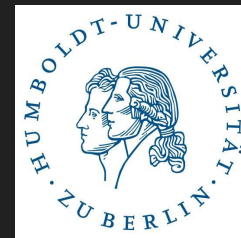
"The HEADT Centre is an independent research organization that belongs to Humboldt-Universität zu Berlin. Elsevier has provided generous funding for the Centre, but does not specify or profit from its research except in so far as the research applies to the scholarly community broadly. Academic freedom is a core value of the Centre.

The HEADT Centre aims to become a major center for research on text and data. It is methodologically broad and includes all ways of managing, manipulating, mining, reading, and understanding text and data in all forms and formats. It will address issues from a variety of stake-holders."

Overview

{headtcentre}

humboldt-elsevier
advanced data & text centre



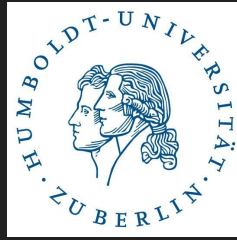
Three topics:

- Vision statement
- Last report
- Survey

Vision Statement 1

{headtcentre}

humboldt-elsevier
advanced data & text centre



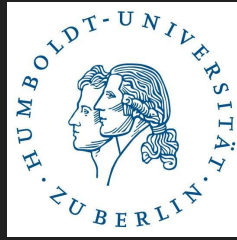
"The near term focus is to become a national and international resource for research integrity in all of its aspects including plagiarism, data falsification, and image manipulation. Discovery mechanisms, replication studies, behavioral aspects and legal considerations all play a role.

This near term focus need not limit future directions."

Vision Statement 2

{headtcentre}

humboldt-elsevier
advanced data & text centre

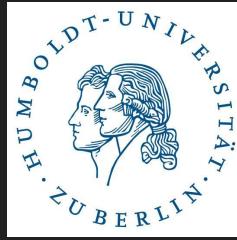


"The HEADT Centre aims to add younger scholars and additional projects to complement and expand its current focus. For example, behavioral and ethnographic research would contribute to understanding research integrity in its social and intellectual environment, and computer-based text and data mining projects could facilitate uncovering information in a wide range of fields including medicine."

Vision Statement 3

{headtcentre}

humboldt-elsevier
advanced data & text centre



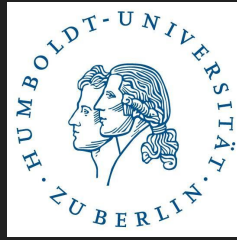
"The research integrity team worked as a service bureau to analyze a complex and highly political case of plagiarism that came to the Kommission Wissenschaftliches Fehlverhalten at Humboldt-Universität zu Berlin.

Part of the work involved the development of a "gray scale" standard that could be used for other cases involving plagiarism. One of the longer term goals is for the HEADT Centre to become a service bureau for the universities in Berlin and potentially Brandenburg."

Report Nov. 2016

{headtcentre}

humboldt-elsevier
advanced data & text centre



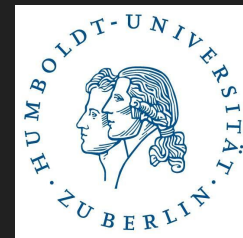
"Research Integrity staff conducted a survey about gray scale issues in plagiarism at the Association of Computer Linguistics Conference in Berlin. A paper analyzing the results should appear on the HEADT Centre website in the foreseeable future as part of an occasional papers series."

The weaknesses of the survey should be clear: it was a convenience sample of people at the Computational Linguistics Conference in August 2016. Roughly 61 people took part. It would be wrong to try to generalize the results beyond this small set of participants. Nonetheless the results were interesting because they reflected the complexity of the issues.

Report Nov. 2016

{headtcentre}

humboldt-elsevier
advanced data & text centre



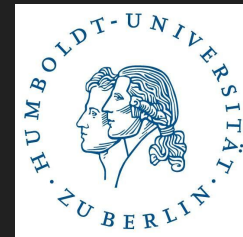
"The Research Integrity group is planning to set up a for-cost certificate program that offers courses on research integrity and may broaden the scope to fit with existing programs on digital curation issues. University College Dublin has expressed interest in a partnership. A partnership with King's College London is also possible.

The program will probably be distance education program with optional on-site visits on the model of the Berlin School's postgraduate executive education."

Survey

{headtcentre}

humboldt-elsevier
advanced data & text centre



PUTATIVE SOURCE: "Accordingly we need to re-examine and re-evaluate the old forms of the book and the old practices of reading and annotation and identify properties that support readers ideal reading experiences.

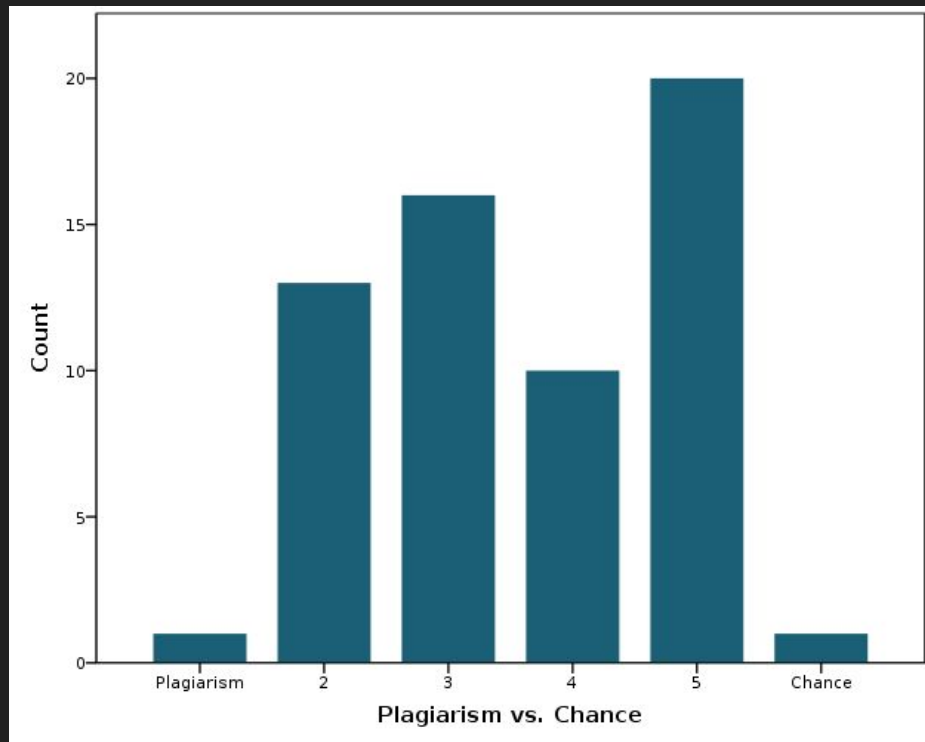
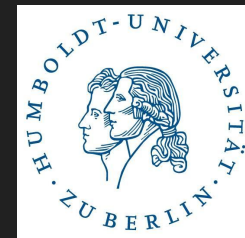
TEXT BEING EXAMINED: "...it can be fruitful to re-examine and re-evaluate older forms of the book and older reading practices. The first paper briefly surveys..."

This text has a limited overlap with the putative source. Out of a sentence of 30 words, eight are the same in two groups, one of two words and one of six.

Survey - 2

{headtcentre}

humboldt-elsevier
advanced data & text centre



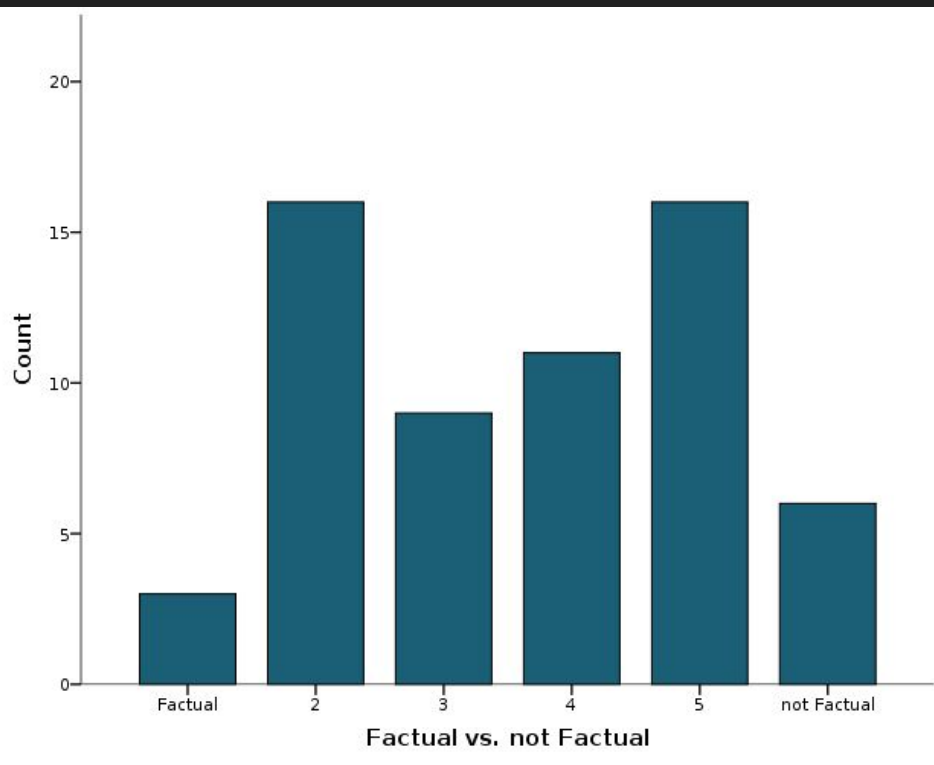
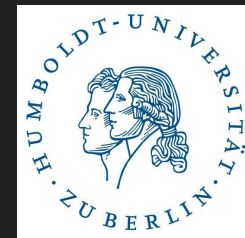
QUESTION: How certain are you that this is plagiarism versus chance?

These results are interesting because of the clear lack of consensus about whether plagiarism occurred or not. For many plagiarism hunters the case would be clear.

Survey - 3

{headtcentre}

humboldt-elsevier
advanced data & text centre



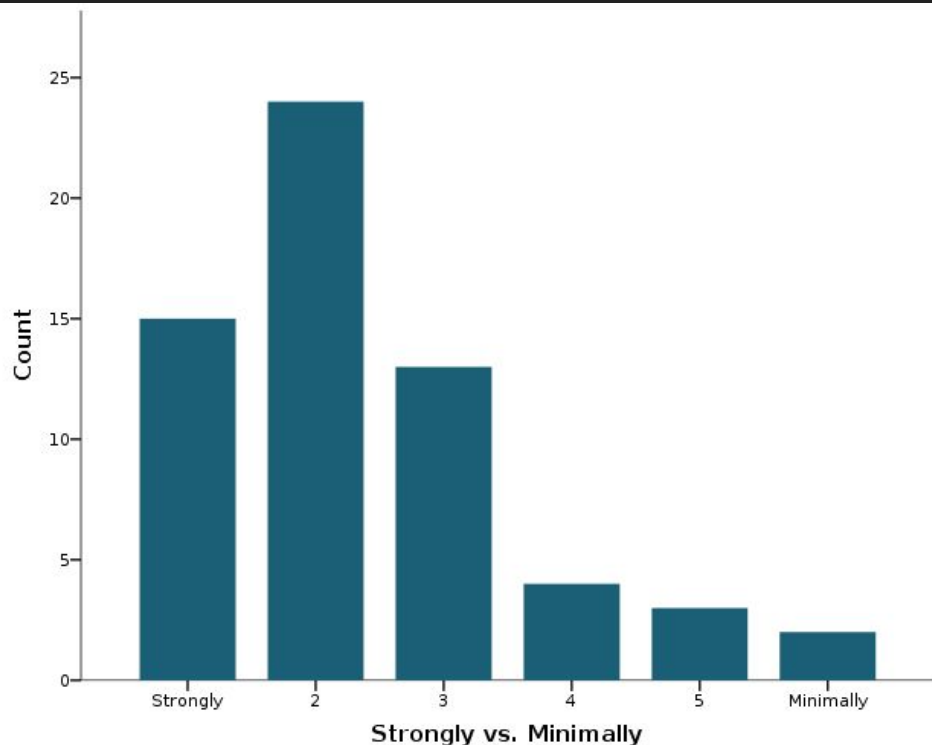
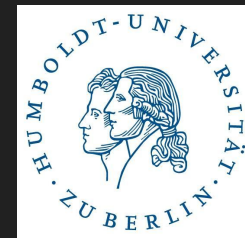
QUESTION: Do you regard the content as primarily factual or not?

The reaction to this question gave a similar split, but with more in each group having strong opinions one way or the other. 28 people tended to regard the text as factual, 33 thought not.

Survey - 4

{headtcentre}

humboldt-elsevier
advanced data & text centre



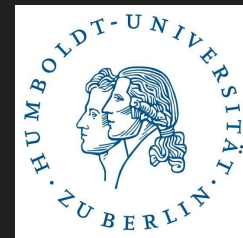
QUESTION: If you knew that other passages in the article came from the same source, would that affect your judgment about this particular passage?

Knowing that other passages came from the same source influenced the judgment of the majority of participants. This suggests that the cumulative impression affects judgment about any one passage

Credits

{headtcentre}

humboldt-elsevier
advanced data & text centre



- *Prof. Johann-Christoph Freytag was co-author of the Vision Statement and the Report.*
- *Stephanie van de Sandt is co-author on the survey analysis. Heinz-Alexander Fütterer created the tables for us.*