

# Investigation of news websites regarding their hyperlinks to external content

Esther von der Weiden, University of Duisburg-Essen

## Introduction

News is, according to Schwiesau (2016), information about current events that are important and interesting for the public. They therefore help to shape public discourse and provide people with a basis for understanding the world. There are many ways for media professionals to make news accessible to the public. In addition to the 'traditional media' such as newspapers, radio and television, the Internet offers a wide range of options. News portals publish the latest news and information on the Internet. The Internet combines many features of other media and therefore has great potential for consumers and media professionals. For instance, current news can be consumed on the Internet around the clock. It is also possible to embed images, videos and audio files in news reports. Another key feature is that it is possible to link to additional content and sources. Such links offer great potential for news portals. For example, readers can be provided with additional information on a particularly interesting topic.

## Research Question

*To what extent do the selected news portals differ from each other in terms of external links?*

The aim of this work is to analyse news portals regarding their links to external content. External links are links that refer to content with a different domain, in contrast to internal links, which refer to content with the same domain. Only links that are part of the news item are analysed. Portals are selected based on statistics on the number of visits to the portals. Therefore, portals that are accessed frequently are primarily analysed. In the process, care is taken to ensure that as diverse a selection of portals as possible is analysed.

## Method

A web crawler in Python is used to collect the data from the news portals and retrieve the individual articles. The framework used for this is Scrapy. The collection is limited to news in text form. Furthermore, only those articles are crawled that can be found on the homepage of the respective news portals. To analyse the stored data, the text of the article is first extracted. The length of the news is then determined based on the number of characters.

## Preliminary Findings

In the preliminary evaluation, two portals are presented as examples: tagesschau.de and zeit.de. As illustrated in the accompanying graphics, it has become evident that these news portals seldom provide external links. With an average external linking rate of 0.94 links per 1000 characters in tagesschau.de and 0.99 links per 1000 characters in zeit.de, the two portals show minimal difference from each other.

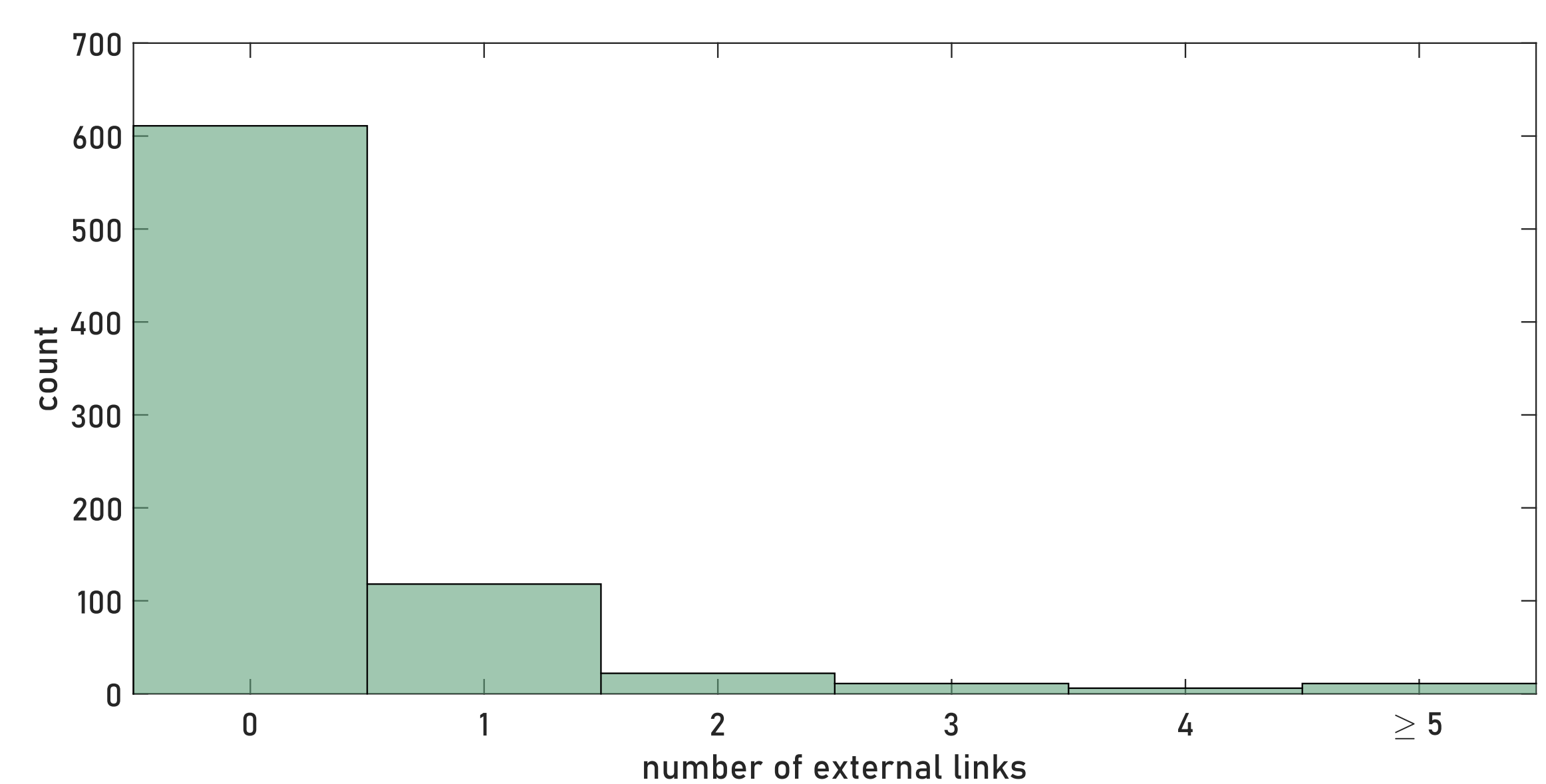


Figure 1: Total count of external links found in articles from tagesschau.de

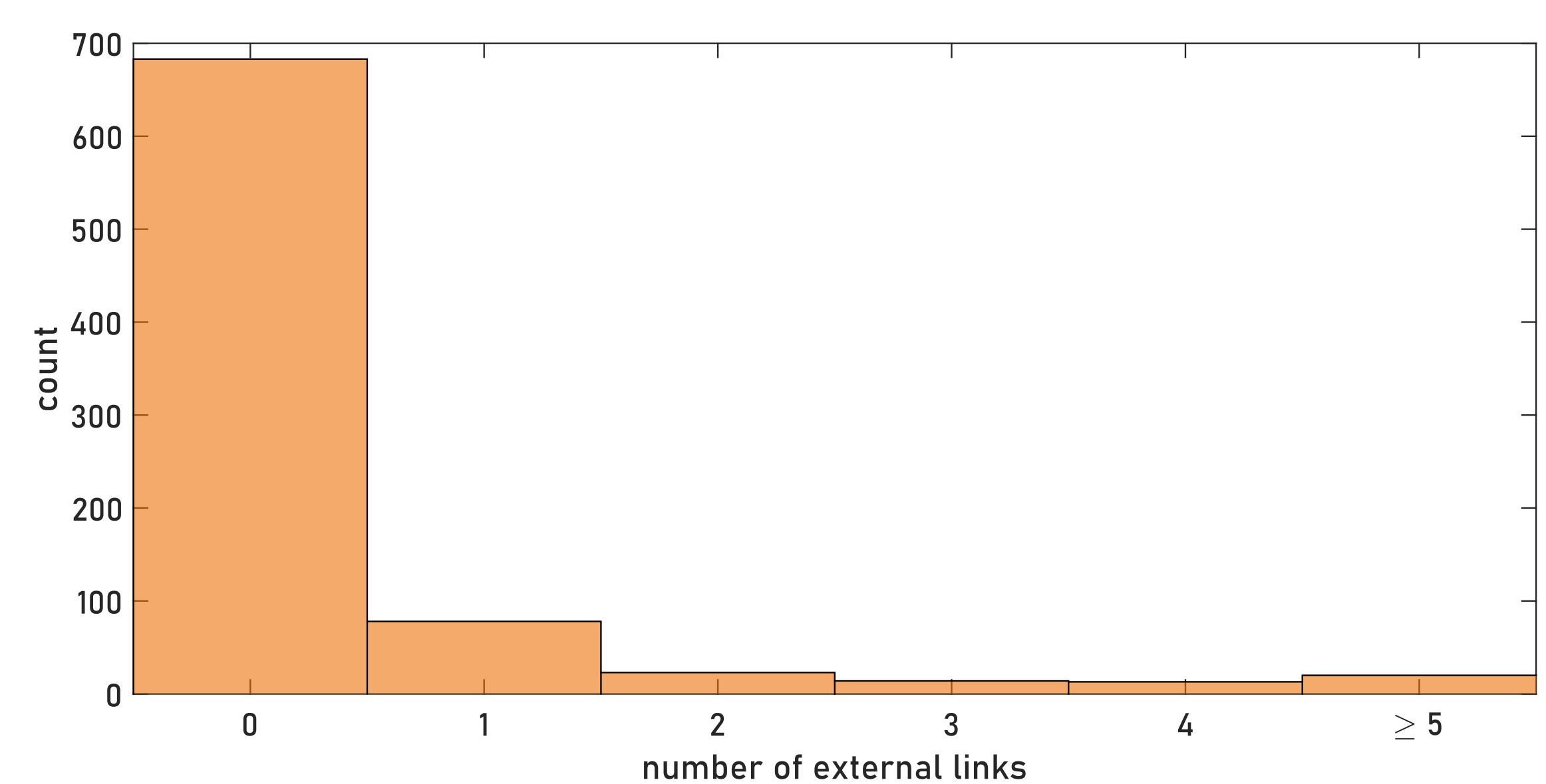


Figure 2: Total count of external links found in articles from zeit.de

## Next Steps

The collected data must be fully analysed. A significant question that arises is whether there is a notable difference between the various news portals. It would be beneficial to ascertain whether there is a correlation between the length of the articles and the number of external links. Determining whether there are notable discrepancies between the individual departments of the respective portals would also be interesting. Furthermore, analysing the domains that are referenced and the frequency with which they are referenced would provide valuable insights.

## References:

1. Schwiesau, D. (2016). Nachrichten - klassisch und multimedial: Ein Handbuch für Ausbildung und Praxis. Springer Fachmedien Wiesbaden; Imprint: Springer VS. <https://doi.org/10.1007/978-3-658-08717-3>