

Kolmogorov Complexity for multidisciplinary domains

João S. Resende, Marco Almeida, Rolando Martins and Luís Antunes

Introduction

Normalized Information Distance (NID) [1] is based on Kolmogorov Complexity and thus uncomputable. It is a universal distance measure for objects of all kinds but for that, it must be approximated based on compression algorithms. We will give a brief overview of the main domains and formulas used.

Normalized Information Distance

It gives minimal information between a sequence X and Y and it can be measured using:

- Normalized Compression Distance (NCD);
- Compression-Based Dissimilarity Measure (CDM)
- Lempel–Ziv Jaccard Distance (LZJD);
- Normalized Relative Compression (NRC);

Public libraries

There are multiple implementations to apply this concept:

- NCD
 - Complearn <https://complearn.org/>
 - Ncdlib <https://github.com/marcoalmeida/ncdlib>
- LZJD
 - jLZJD <https://github.com/EdwardRaff/jLZJD>

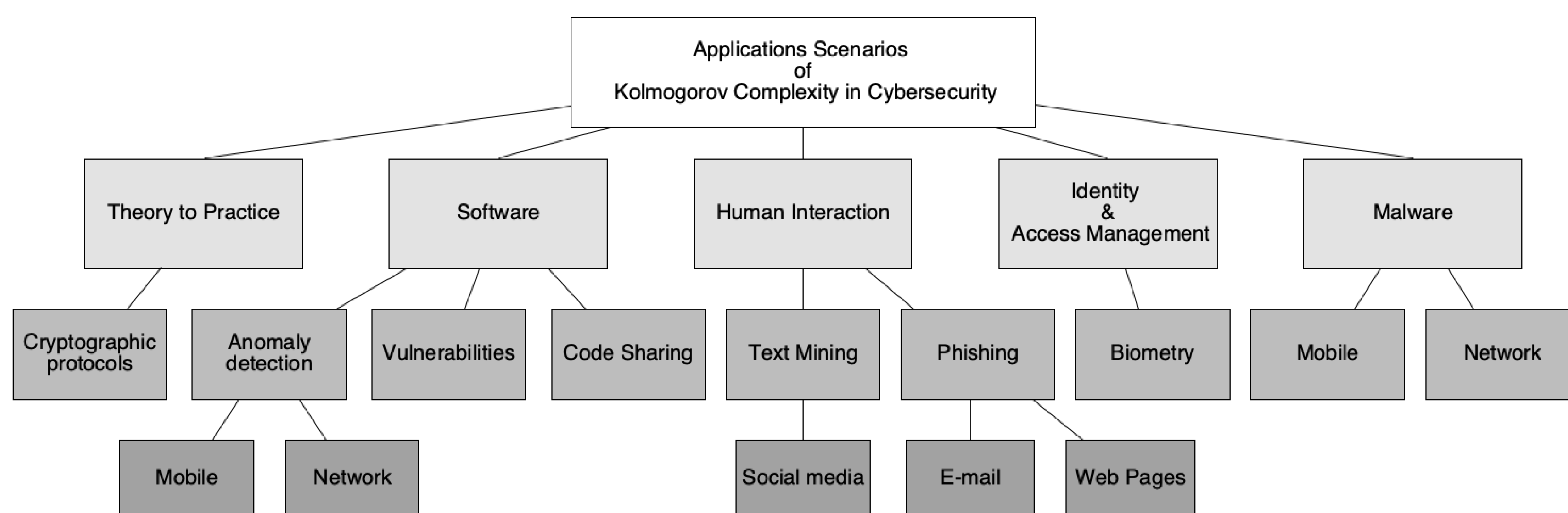


Fig.1. Taxonomy in cybersecurity [2]

Cybersecurity (an example)

- Figure 1 demonstrates the multiple domains of cybersecurity where NID was used;
- Many articles in this domain are from malware classification and human interaction (with natural language processing);
- The most recent paper focus on evaluating cryptographic protocols [3].

Conclusions

- NCD contains multiple applications domains;
- The high number of publications with NCD is from Computer Science
 - Specially in Cybersecurity [2];
- From our review, we found five main application domains (Figure 2)

NCD

- NCD is the most used formula that uses NID;
- In Figure 2, we have the most common application domains of NCD;
- Many articles implement NCD, but this study focus only on the ones that cited the original publication.

Acknowledgements

This work of João S. Resende was supported by Fundação para a Ciência e Tecnologia (FCT), Portugal (PD/BD/128149/2016). This work has been supported by the EU H2020-SU-ICT-03-2018 Project No. 830929 CyberSec4Europe (cybersec4europe.eu).

References

- [1] Vitányi, Paul MB, et al. "Normalized information distance." Information theory and statistical learning. Springer, Boston, MA, 2009. 45-82.
- [2] S Resende, João, Rolando Martins, and Luís Antunes. "A Survey on Using Kolmogorov Complexity in Cybersecurity." Entropy 21.12 (2019): 1196.
- [3] Resende, João S., et al. "Breaking MPC implementations through compression." International Journal of Information Security 18.4 (2019): 505-518.

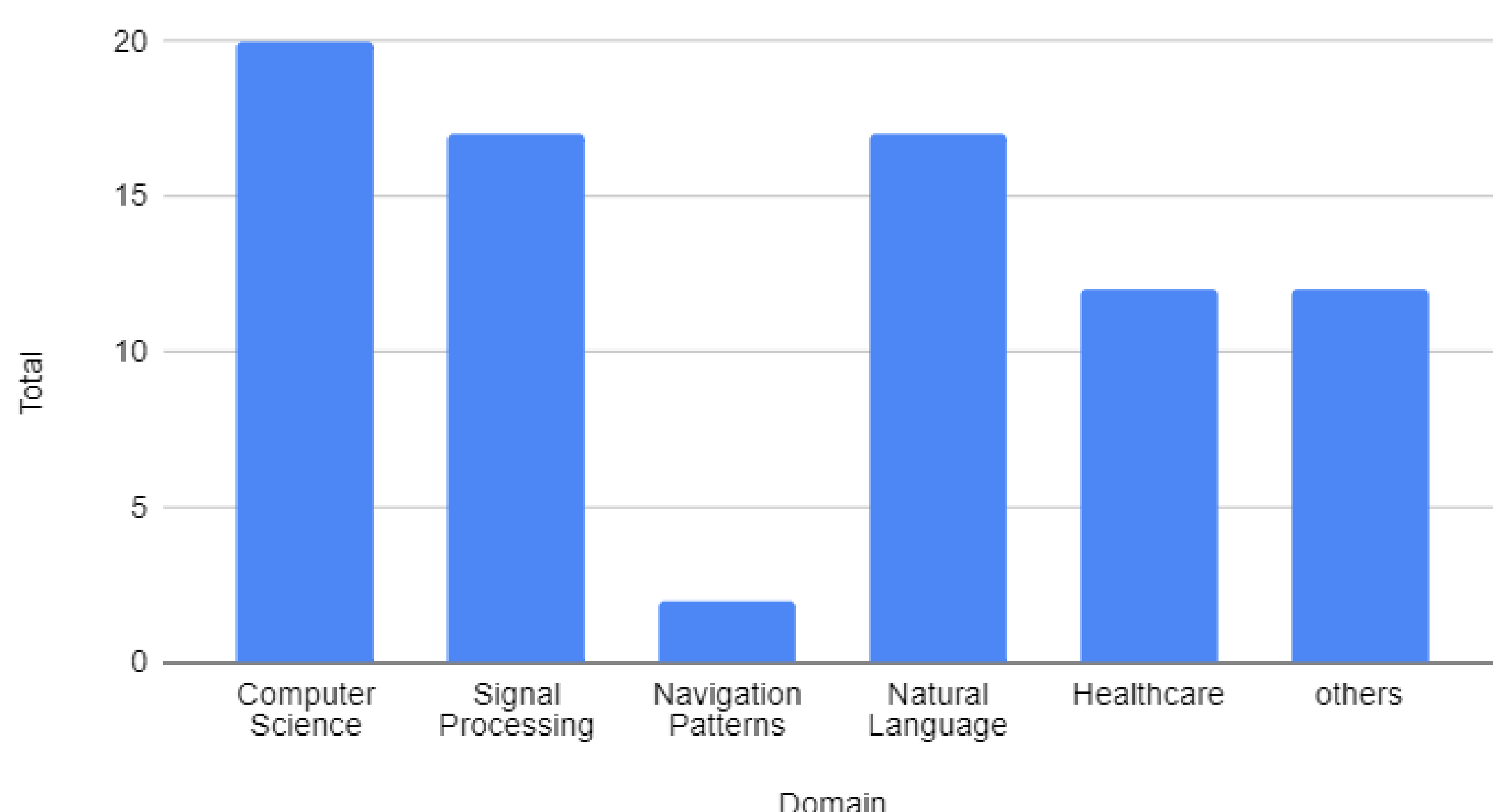


Fig.2. Number of articles by domain