I will talk about algorithms that find and eliminate redundant information in representations of data. The solutions I present are heavily inspired by classical information theory and modern machine learning. Relevant papers are:

[1] Compressing multisets with large alphabets
D Severo, J Townsend, A Khisti, A Makhzani, K Ullrich - 2022 Data Compression Conference (DCC), 2022

[2] Lossy compression for lossless prediction