TF–IDF for terms of interest in abstracts (top 5000 terms)