Supplementary File 1

Table 1: Overview of the main type of information contained in the dataset. The full codebook is available on the following Gitlab repository https://gitlab.com/shoricitza/afsj-pol-lex-track-quantitative-dataset

Information about	Type of information
Legal act	Interinstitutional code
	Title of the legislative procedure
	Type of procedure
	Type of act
	Number of readings
	Stage reached during the legislative procedure
	Legal basis
	Directory codes 19.01 General 19.10 Free movement of persons - 19.10.10 Elimination of internal border controls - 19.10.20 Crossing external borders - 19.10.30 Asylum policy - 19.10.40 Immigration and the right of nationals of third countries 19.20 Judicial cooperation in civil matters 19.30 Police and judicial cooperation in criminal and customs matters 19.40 Programmes 19.50 External relations
	Total duration of the procedure in number of days
	Signature by the President of the EP and Council date; format DD/MM/YY
European Parliament	Name and party of the rapporteur
	Name and party of the co-rapporteur
	Date of the committee opinion; format DD/MM/YY
	Committee responsible and committee for opinion
	Date of the committee referral; format DD/MM/YY
	Name and party of the rapporteur for opinion
	Date of the EP's opinion/ position at 1st, 2nd and 3rd reading; format DD/MM/YY
	Date of the EP's formal reconsultation; format DD/MM/YY
	Decision type by Parliament, 1st, 2nd and 3rd reading.
	Number of votes in favor, against, abstentions (EP)
	Total number of tabled and adopted amendments (EP)
Council	Date of the adoption of the legislative act by the Council; format DD/MM/YY
	Number of points A on the agenda of the Council
	Number of points B on the agenda of the Council
	Act adopted by the Council after consultation of the Parliament date; format DD/MM/YY
	Political agreement in the Council date; format DD/MM/YY
	Approval by the Council of the legislative act at 1st , 2nd , 3rd reading date; format DD/MM/YY
	Position of the Council at 1st, 2nd reading, 3rd reading date; format DD/MM/YY
Commission	Date of the adoption by Commission; format DD/MM/YY
	Date of the corrigendum; format DD/MM/YY
	Date of the transmission of the legislative draft to the EP and the Council; format DD/MM/YY
	Commission's position on EP amendments at each reading
	Withdraw by the Commission date; format DD/MM/YY

Presentation of the FuzzyWuzzy Package in Python

FuzzyWuzzy computes the standard Levenshtein distance similarity ratio between two sequences. The formal definition of the Levenshtein distance between two strings a and b can be seen as follows:

$$\mathrm{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \mathrm{lev}_{a,b}(i-1,j) + 1 \\ \mathrm{lev}_{a,b}(i,j-1) + 1 \\ \mathrm{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

Where 1(ai≠bj) is 0 when a=b and 1 when a≠b. The rows on the minimum above correspond to a deletion, an insertion, and a substitution in that order. However, contrary to the Levenshtein distance, which is sensitive to the order of the words in the text, stop words etc., FuzzyWuzzy offers a more accurate way to determine how similar two strings are. It has more powerful functions that allow to deal with substring matching, different word order, lower/upper cases, punctuation, 'stopwords', such as 'the', 'from' etc. Though other similarity algorithms, such as Jaro-Winkler, Jaccard Index, Sorensen-Dice Index, or Ratcliff-Obershelp, have been tested, the FuzzyWuzzy Package performed better.