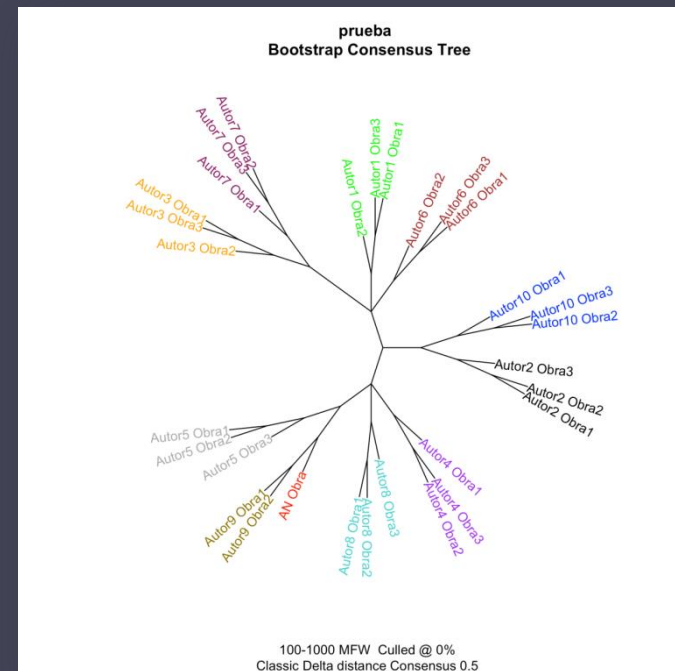


Profundización en stylo. Verificación de autoría



Laura Hernández Lorenzo

Humanidades Digitales. Del corpus a la interpretación:
Estilometría con R (Universidad de Burgos)

Contenidos

- Atribución / verificación de autoría
- Verificación de autoría
- General Imposters Framework
- Práctica 6

Métodos de atribución de autoría

- Se intenta encontrar el autor de un texto anónimo dentro de un conjunto de candidatos
- Es decir:
 - *De todos estos autores, ¿quién es el autor más probable de la obra anónima?*
 - *De todos estos autores, ¿cuál es el más cercano estilísticamente al texto anónimo?*

Métodos de atribución de autoría

- Problema: ¿y si el autor del texto anónimo no está entre los candidatos que se han recogido para el análisis?
 - Problema del vecino más cercano → siempre habrá un candidato estilísticamente más cercano al texto anónimo

Métodos de verificación de autoría

- Se trata de confirmar o rechazar si el principal candidato o sospechoso de haber escrito el texto anónimo es efectivamente el autor
- Se supera el problema del ‘vecino más cercano’

Métodos de verificación de autoría

- Destaca el marco de los Impostores -‘General Imposters (GI) framework’-
 - Introducido por primera vez en Estilometría y problemas de autoría por Koppel y Winter (2014).

Determining If Two Documents Are Written by the Same Author

Moshe Koppel and Yaron Winter

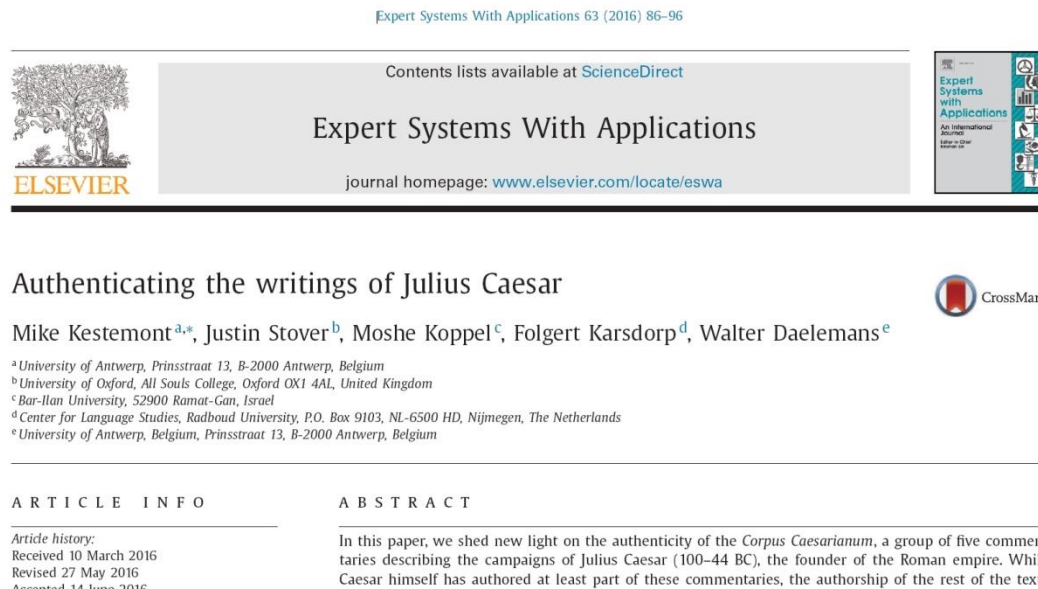
Department of Computer Science, Bar-Ilan University, Ramat-Gan, Israel 52900.
E-mail: {moishk, yaron.winter}@gmail.com

Almost any conceivable authorship attribution problem can be reduced to one fundamental problem: whether a pair of (possibly short) documents were written by the same author. In this article, we offer an (almost) unsupervised method for solving this problem with surprisingly high accuracy. The main idea is to use repeated feature subsampling methods to determine if one document of the pair allows us to select the other from among a background set of “impostors” in a sufficiently robust manner.

problem, regardless of the number of candidates. All we need to do is ask if the anonymous text was written by each of the respective candidates; we will get a positive answer for the true author and a negative answer for all the others. On the other hand, the verification problem is strictly harder than the attribution problem: the fact that we solve a closed-set attribution problem offers no guarantees that we can solve an open-set verification problem. It is thus not surprising that, with a single limited exception (see below), no satisfactory solution has

Impostores

- En estudios literarios, destaca su aplicación a los problemas autoriales con las obras de Julio César (Kestemont et al., 2016).



Impostores

- Implementado en mayo de 2018 en el paquete ‘stylo’ en R de Eder, Rybicki y Kestemont
 - <https://computationalstylistics.github.io/blog/imposters/>
- Valores resultantes entre 0 y 1
 - **0** significa que con toda seguridad **no** es el autor
 - **1** significa que con toda seguridad **sí** es el autor

Práctica 6

- Abre en RStudio el script “taller7.R”
- ¡Manos a la obra!