



Modelos de Recuperación de Información (Parte 2)

Fecha entrega: 29/04/2022

Para la entrega del TP resuelto arme un único archivo comprimido (.tar.gz) y envíelo a través del siguiente formulario: <https://forms.gle/nMMrV2XghDVb4zR19> el cual se encontrará habilitado hasta la fecha de entrega establecida.

Bibliografía sugerida: MIR [1] Capítulos 2, MAN [2] Capítulos 1,7,12.



1. Retome el TP de "Modelos de RI" y calcule el modelo de lenguaje (unigramas) para los documentos del ejercicio 2. Utilizando el modelo de Query Likelihood calcule los rankings para las siguientes consultas:

- a) país cultura
- b) país libre cultura
- c) software propietario licencia

¿Qué problemas encuentra? Luego, calcule las probabilidades de los términos utilizando una combinación con el ML de la colección (suavizado Jelinek-Mercer, $\lambda = 0,7$). Compare con las probabilidades anteriores y explique las diferencias. Repita las consultas con los nuevos valores. Explique los resultados.

2. Repita el ejercicio pero esta vez utilice la divergencia de Kullback-Leiber y un suavizado por Dirichlet-Priors utilizando para los parámetros los valores sugeridos en la literatura.
3. Utilizando modelos de lenguaje en Terrier (use Hiemstra LM), repita los experimentos del ejercicio 9 del TP de "modelos" y compare los resultados con los anteriores. ¿Son consistentes? Calcule las métricas apropiadas para comparar los diferentes sistemas y configuraciones.

Referencias

- [1] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval: The Concepts and Technology Behind Search*. Addison-Wesley Publishing Company, USA, 2nd edition, 2008.
- [2] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.