## Reporte sobre la Extracción de Entidades

Elías González Nieto

6 de junio de 2025

## Extracción de Entidades: Resultados Principales

Durante el proceso de extracción de entidades en los libros de texto gratuitos de la SEP, se obtuvieron los siguientes resultados relevantes:

- Número total de entidades: Se detectaron un total de 3748 entidades, considerando todas las apariciones, incluidas las duplicadas. De estas, 1784 fueron entidades únicas, lo cual indica que hubo un número significativo de repeticiones entre los textos analizados.
- 2. Entidades más frecuentes: A continuación se muestra una visualización con las entidades que tuvieron más de 50 apariciones. Estas entidades reflejan los personajes, lugares o conceptos más presentes en los libros de historia.

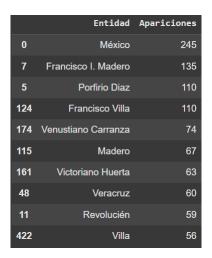


Figura 1: Entidades más frecuentes

3. Entidades duplicadas por libro y generación: La siguiente tabla muestra el número de entidades duplicadas identificadas en cada uno de los libros y generaciones analizadas.

Libro	Total	Duplicados	Únicos	% Duplicados
196012	62	14	48	22.58%
1960l3	527	66	461	12.52%
196211	76	12	64	15.79%
196212	550	77	473	14.00%
197211	32	6	26	18.75%
198211	39	12	27	30.77%
199311	444	49	395	11.04%
1993l2	563	65	498	11.55%
200811	443	53	390	11.96%
200812	320	27	293	8.44%
201111	363	33	330	9.09%
2014l1	448	33	415	7.37%

Cuadro 1: Resumen de entidades duplicadas por libro y generación