

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:

1. **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.

	Entidad	Apariciones
0	México	245
7	Francisco I. Madero	135
5	Porfirio Díaz	110
124	Francisco Villa	110
174	Venustiano Carranza	74
115	Madero	67
161	Victoriano Huerta	63
48	Veracruz	60
11	Revolución	59
422	Villa	56

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
1960l2	62	14	48	22.58 %
1960l3	527	66	461	12.52 %
1962l1	76	12	64	15.79 %
1962l2	550	77	473	14.00 %
1972l1	32	6	26	18.75 %
1982l1	39	12	27	30.77 %
1993l1	444	49	395	11.04 %
1993l2	563	65	498	11.55 %
2008l1	443	53	390	11.96 %
2008l2	320	27	293	8.44 %
2011l1	363	33	330	9.09 %
2014l1	448	33	415	7.37 %

Cuadro 1: Resumen de entidades duplicadas por libro y generación