

TORREÓN EN EL MACROSCOPIO

Historia cuantitativa de problemas sociales en cien
años de producción noticiosa de *El Siglo de Torreón*

Sergio Garza Orellana

Problema social

Constructo social

Recursos públicos

Congreso



Arenas públicas

Periódico



Atención pública

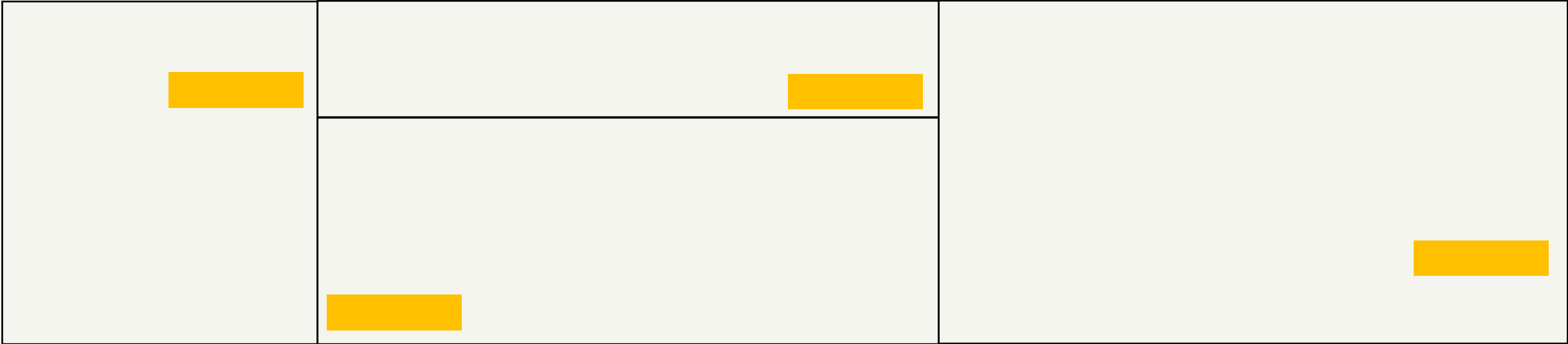
Televisión

[illegible]

Academia

Problema social
Constructo social

Recursos públicos **Arenas públicas** **Atención pública**
Congreso Periódico Televisión



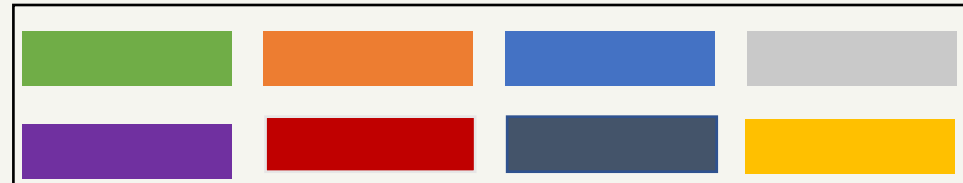
Academia

¿Cómo identificarlos?

Problema social

Constructo social

Periódico



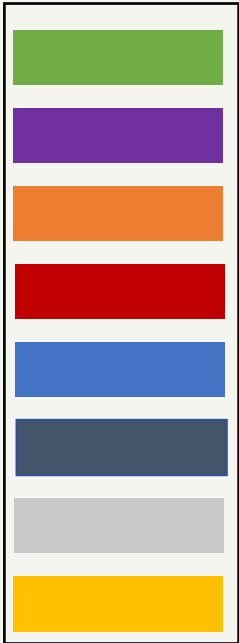
¿Cómo identificarlos?

Problema social

Constructo social

Periódico

1922



¿Cómo identificarlos?

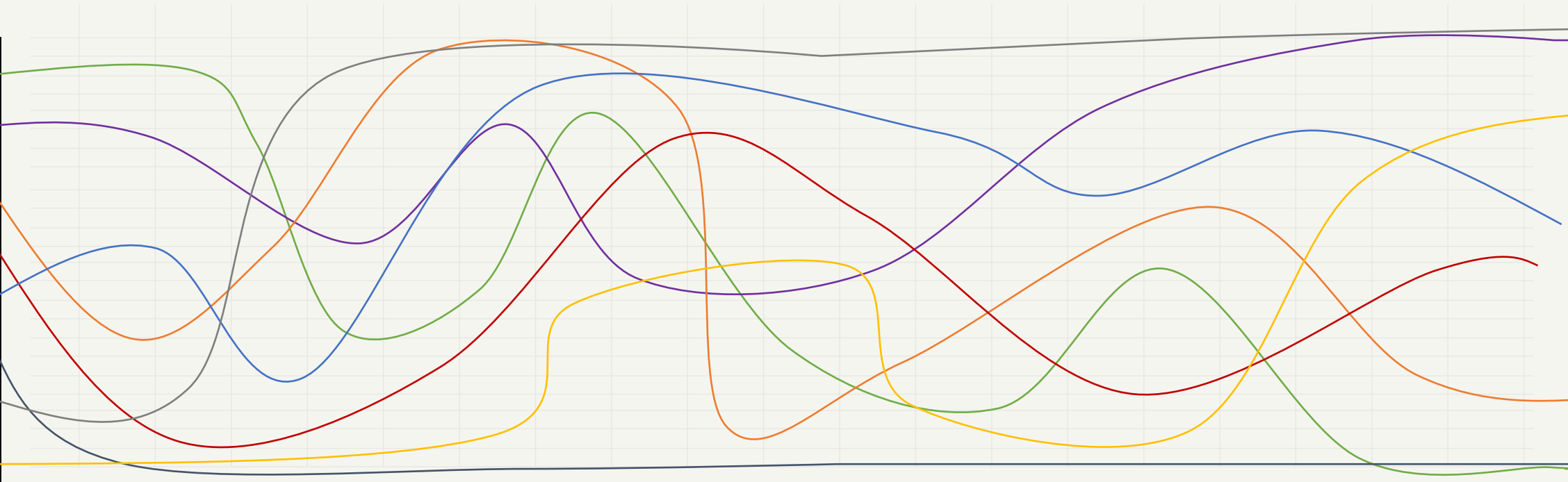
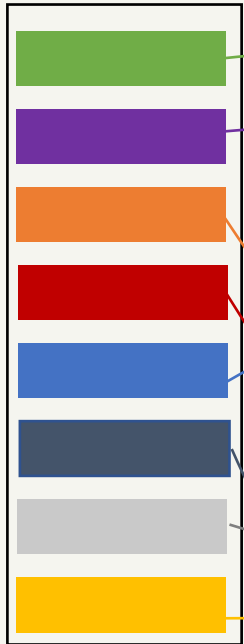
Problema social

Constructo social

¿Cómo se han transformado?

Periódico
1922

1927 1932 1937 1942 1947 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 2002 2007 2012 2017 2022





El Siglo de Torreón

PERIÓDICO REGIONAL • DEFENSOR DE LA COMUNIDAD

300
notas diarias

Python / Servidor local

**¿Cómo leer
11,000,000
de notas?**

Problemas sociales

Modelado de tópicos
[*Topic Modelling*]

Entidades nombradas
[*Named entities*]



Análisis de redes



Visualización



INTERPRETACIÓN HISTÓRICA

Topics

gene 0.04
dna 0.02
genetic 0.01
...

life 0.02
evolve 0.01
organism 0.01
...

brain 0.04
neuron 0.02
nerve 0.01
...

data 0.02
number 0.02
computer 0.01
...

Documents

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions

"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

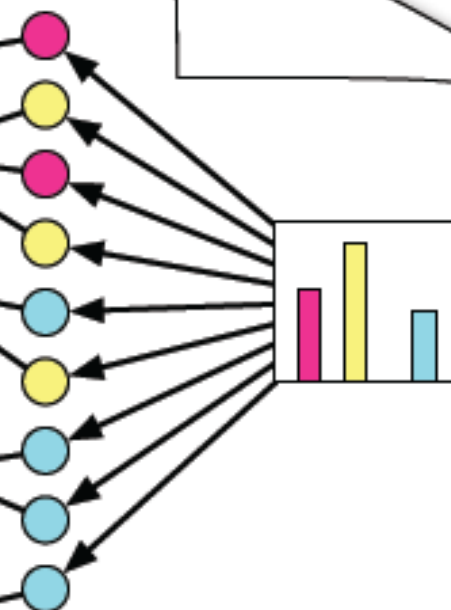


* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

Topic proportions and assignments



OBJETIVO GENERAL

Identificar las principales **problemáticas sociales** abordadas por *El Siglo de Torreón* y **analizar su transformación histórica**, mediante el diseño, desarrollo, implementación y evaluación de métodos computacionales procesamiento de lenguaje natural en el corpus del periódico de 1922 a 2022.

OBJETIVOS ESPECÍFICOS

- 1) **Sistematizar y ordenar una base de datos** general con la información digitalizada del diario *El Siglo de Torreón*, desde 1922 a la actualidad.
- 2) **Desarrollar y adaptar métodos de procesamiento de lenguaje natural** que permitan la identificación de tópicos centrales y el reconocimiento de entidades nombradas, para la construcción de la variable [Problema social]
- 3) **Analizar diacrónicamente las transformaciones** de los problemas sociales a lo largo del periodo de tiempo analizado, a través de técnicas estadísticas y de análisis de redes.
- 4) **Crear visualizaciones** de la información anterior, que permitan la lectura, identificación e interpretación de procesos sociales históricos durante el periodo de tiempo analizado.
- 5) **Evaluar los resultados** anteriores a través de una comparativa con otros estudios históricos de la región durante el periodo analizado, con el fin de establecer similitudes y diferencias de interpretación entre esta metodología de investigación histórica y otras.
- 6) **Generar conclusiones**, definir límites y prever alcances de este sistema de análisis para otras fuentes hemerográficas con características similares a la del objeto de estudio.

ETAPA 1
**Construcción de
base de datos**

**Base de datos
relacional en MySQL**

**SEGMENTACIÓN
2001 - 2021:
Webscraping**

**1922 - 2001:
Análisis de grado de
reconocimiento OCR**

**Colaboración con
*El Siglo de Torreón***

Python / Servidor local

ETAPA 2
**Procesamiento de
lenguaje natural**

SEGMENTACIÓN

Modelado de tópicos

**Reconocimiento de
entidades nombradas**

**Aumento en base de
datos MySQL**

Python / Servidor local

ETAPA 3
**Análisis diacrónico
y reticular**

**ANÁLISIS DE LA
INFORMACIÓN**

**Redes
Frecuencias
Patrones
Transformaciones
dentro de tópicos**

Python / Servidor local

ETAPA 4
**Visualización de
los datos**

**Publicación en
web**

Django / HTML

**Colaboración con
departamento de
sistemas UIA Torreón**

**Dominio y
alojamiento web**

ETAPA 5
**Evaluación de
resultados**

**Cotejo con
historiadores locales**

**Colaboración con
Archivo Histórico Ibero
Torreón y Municipal**

ETAPA 6
**Análisis de
replicabilidad y
publicación de
resultados**

TORREÓN EN EL MACROSCOPIO

Historia cuantitativa de problemas sociales en cien
años de producción noticiosa de *El Siglo de Torreón*

Sergio Garza Orellana