

# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## MODELADO DE TÓPICOS CON LDA

El modelado de tópicos es un conjunto de herramientas estadísticas e informáticas que permiten identificar de manera rápida y automatizada los tópicos latentes o temas subyacentes en un conjunto numeroso de textos de interés (por ejemplo, las respuestas a campos abiertos en una consulta pública de interés para el Estado, las noticias publicadas en portales digitales en un día, etc).

LDA es una técnica para el modelado automatizado de tópicos que ayuda a identificar los tópicos en un conjunto numeroso de textos que tengan características similares en forma y contenidos.

## ¿QUÉ ES UN TÓPICO?

Un tópico es un conjunto de palabras que tienden a aparecer juntas en un determinado conjunto de documentos.

Por ejemplo, en un grupo de artículos científicos es esperable encontrar tópicos relativos a las distintas disciplinas científicas (química, matemática, biología, física,, etc), así como tópicos relativos al discurso científico en general (metodología, protocolos, instrumentos, mediciones, etc)..

Sería mucho menos probable encontrar tópicos relativos a la moda, por ejemplo, que sí serían comunes en un conjunto de documentos conformado por revistas de variedades.



# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## ¿PARA QUÉ SIRVE EL MODELADO DE TÓPICOS?

El modelado de tópicos ayuda a leer de manera rápida y automatizada los temas que subyacen a un conjunto de textos que compartan características formales y/o de contenidos (corpus de análisis). En el caso de una consulta abierta, este conjunto de textos estaría constituido por el total de respuestas de las y los participantes en los campos abiertos de la consulta. Esta herramienta también puede ser aplicada al análisis de tendencias y matrices de opinión en medios digitales, así como al análisis de tópicos de interés en publicaciones científicas, o en discursos de naturaleza política.

## POTENCIALES USOS

### Estudio de matrices discursivas

Pueden ser útiles para el estudio de conjuntos de documentos como titulares de prensa y reportajes. Por lo tanto, puede contribuir con el estudio de matrices mediáticas o comunicacionales. En esta categoría, puede ser que el corpus se encuentre integrado por otro tipo de textos, como discursos políticos, entonces sería posible estudiar un buen número de alocuciones de una cierta tendencia para extraer los aspectos de significado más constantes.

### Investigación documental

Puede emplearse para el análisis de corpus de productos de investigación, tales como libros y artículos científicos. La importancia de los repositorios de publicaciones gubernamentales y universitarias podría incrementarse si contáramos con buscadores que no sólo extrajeran los títulos de los artículos, sino que además los analizaran entregándonos como resultado un conjunto de tópicos relevantes en función de los temas de interés en campos como la salud y la producción alimentaria.

### Diseño y evaluación de políticas públicas

El modelado de tópicos puede contribuir como herramienta de organización de documentos recogidos a partir de consultas públicas en las que los ciudadanos expresan sus opiniones en torno a varios temas. Por ejemplo, una consulta pública sobre una iniciativa legislativa o del Ejecutivo podría registrarse y analizarse eficientemente con ayuda de esta herramienta, además de que los textos pueden colocarse a disposición del público. Lo mismo sería válido para diferentes tipos de consulta a nivel nacional,

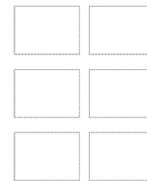
# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## ¿CÓMO FUNCIONA EL MODELADO DE TÓPICOS?

### DEFINICIÓN DE CORPUS

1. El modelado de tópicos requiere en primer lugar de la definición del corpus de análisis o conjunto de textos a analizar. Tales textos deben ser adecuados a un mismo formato que permita su procesamiento.

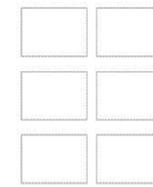
CORPUS



Un corpus es un conjunto de textos que comparten características formales textuales (cartas, notas de prensa, discursos políticos) .

### PREPROCESAMIENTO GRAMATICAL

2. Los textos deben ser procesados usando una herramienta informática que permita identificar las partes gramaticales que conforman los mismos,. Por ejemplo, el *freeling* (herramienta en código abierto desarrollada por Lluís Padró, y mantenida por TALP Research Center en la Universitat Politècnica de Catalunya), que identifica y agrupa las palabras que conforman el corpus según su naturaleza gramatical (verbos, sustantivos, adjetivos, etc)



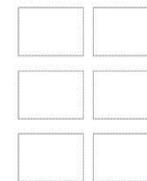
ABC

Análisis gramatical

### PROCESAMIENTO Y MODELADO DE TÓPICOS

3. Seguidamente los textos se procesan usando *LDA*, implementación en C del algoritmo Latent Dirichlet Allocation (LDA), que permite analizar el corpus y extraer los tópicos subyacente que combinados forman el conjunto de documentos.

Modelado de Tópicos



# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## MODELADO DE TÓPICOS DE UN CORPUS MEDIÁTICO

### DEFINICIÓN DE CORPUS

El corpus de análisis se construyó mediante la recopilación automatizada (*web scrapping*) de notas publicadas en portales digitales con amplia repercusión en la agenda mediática venezolana (*La Patilla, Runrunes, Correo del Caroní, El Nacional, Tal Cual, Dólar Today, Informe21, y Noticiero Digital*) entre los días 6 y 8 de marzo de 2016.



# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## MODELADO DE TÓPICOS DE UN CORPUS MEDIÁTICO

### PREPROCESAMIENTO GRAMATICAL

El preprocesamiento gramatical automatizado mediante el freeling permite excluir términos de uso frecuente en el tipo de discurso que caracterice a los textos analizados, con el fin de depurar los tópicos (por ejemplo, en el caso del corpus mediático, se identificaron verbos típicos del discurso periodístico como: declarar, decir, confirmar, afirmar, hablar; así como sustantivos como Venezuela, país, día, año que aparecen en todos los tópicos identificados y fueron depurados del corpus)

**Correo del Caroní**  
La Casa de las Ideas

LABORAL POLÍTICA ECONOMÍA CIUDAD SUCEOS DEPORTES CULTURA  
INTERNACIONAL MÁS CASO CORRO DEL CARONI

Lunes, 07 de junio 2016 09:00

### Minero sobreviviente confirma balacera y muertes en yacimiento de oro en Tumeremo

Clavel A. Rangel Jiménez

Valorar este artículo

Miedo, desolación y ningún rastro de mineros desaparecidos a 48 horas de protesta en Tumeremo

FOTOS, INFOGRAFÍA Y MAPA | 84 muertos en protestas en Venezuela hasta el #5Jun

RUNRUNES NACIONAL ECONOMÍA INTERNACIONAL OPINIÓN RELAX 5+5 TECH

RRES PLUS

### Minero sobreviviente confirma balacera en yacimiento de oro en Tumeremo

Por Clavel Rangel | Fecha: 07/03/2016

Compartir

La Virgen del Valle reclama a sus hijos: 7 margariteños víctimas de la justicia militar fueron enviados a La Pica

La Virgen del Valle reclama a sus hijos: 7 margariteños víctimas de la justicia militar fueron enviados a La Pica

FOTOS, INFOGRAFÍA Y MAPA | 84 muertos en protestas en Venezuela hasta el #5Jun

fotos desde tpic.googleyndication.com...

08 de junio 2017

**Tal Cual**  
claro y raspa

100% Venezolano

OPINIÓN NACIÓN ECONOMÍA EDITORIAL MUNDO OTRAS HISTORIAS

TEMAS DEL DÍA CONSTITUYENTE CHIMBA ORTEGA GÁZ OLLAS VACÍAS

INICIO NACIÓN

### Familiares buscan a mineros desaparecidos y solo consiguen rastros de aceite

lapatilla 08/06/2017 | Actualizado: 11:06 am

Buscar

Camión con cuerpos estaría en la mina "Hoja de Lata", dice sobreviviente de la masacre de Tumeremo

Mar 5, 2016 7:10 am  
Publicado en: Actualidad, Nacionales, Sucesos

Sobreviviente de masacre de Tumeremo implica a cuerpos policiales con "El Topo"

**El Pitazo** CARTUCHO ALIÑO

INICIO GRAN CARACÁS CENTRO GUAYANA LOS ANDES LOS LLANOS OCCIDENTE ORIENTE M

SUCESOS POLÍTICA ECONOMÍA DEPORTES TECNOLOGÍA INTERNACIONAL ESPECIALES

PITAZOS DEL DÍA • Protestas de la oposición • Asamblea Constituyente • Huelgafanos de la salud • Caso Odebrecht • Decreto di

El Pitazo en la calle

El Pitazo en la radio

El Pitazo investiga

Alianzas

### CRÓNICA | Minero sobreviviente confirma balacera y muertes en yacimiento de oro en Tumeremo

El Pitazo | El Pitazo

Marzo 7, 2016 10:38 pm

La versión de otro de los testigos coincide al asegurar que un grupo de mineros fue agrupado en la mina y disparados con ráfagas de tres. Algunos fueron cortados con machete

La tranca de la troncal 10 se mantiene en la entrada de Tumeremo! Foto: Corisía @Playbesa

La tranca de la troncal 10 se mantiene en la entrada de Tumeremo! Foto: Corisía @Playbesa

Por Clavel Rangel Jiménez

La mañana del viernes los mineros viajaron a la zófa del Miamo, en la frontera de los municipios

# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## MODELADO DE TÓPICOS DE UN CORPUS MEDIÁTICO

### PROCESAMIENTO Y MODELADO DE TÓPICOS

La herramienta permite identificar entre 5 y 100 tópicos, de acuerdo con el corpus analizado y las necesidades de análisis del mismo. En este caso se puede apreciar 4 de los 20 tópicos más destacados en el corpus analizado mostrados en la interfaz de listado de tópicos. Estos tópicos ayudan a entender rápidamente los principales temas desarrollados en el conjunto de textos.

Listado de Tópicos

20 Topics ▾

**Tópico #0**  
protesta, agua, protestas, protestan, seguridad, fueron, zona, vecinos, hechos, despues, hora, hecho, familiares, manifestantes, escasez, calles, publica, avenida, salieron, municipio  
Ver Documentos para Tópico #0

**Tópico #1**  
cambio, oficial, agencia, fuerza, defensa, medio, destaco, fueron, putin, ruso, incluso, rusia, rusos, humanos, historia, espera, publico, ayer, diciendo, previsto  
Ver Documentos para Tópico #1

**Tópico #2**  
acuerdo, poder, reuters, delitos, decision, podria, politicos, tendra, ministro, juez, tribunal, estar, cometido, francia, blanqueo, dinero, corte, venezolana, judicial, martinez  
Ver Documentos para Tópico #2

**Tópico #3**  
maduro, ministro, civil, pueblos, vicepresidente, bolivariana, actividades, medidas, nicolas, relaciones, justicia, familia, lopez, obama, aeropuerto, plena, general, diario, expreso, califica  
Ver Documentos para Tópico #3

### Tópico 0

Relativo a **protestas** (*protesta, seguridad, manifestantes, vecinos, salieron, calles*)

### Tópico 1

Relativo a **Rusia** (*Rusia, Putin, agencia, oficial, cambio*)

### Tópico 2

Relativo a **delito de capitales** (*blanqueo, dinero, corte, delitos, cometido, decisión, juez, tribunal*)

### Tópico 3

Relativo a **medidas gubernamentales** (*Maduro, Nicolás, Vicepresidente, medidas, ministro, justicia*)

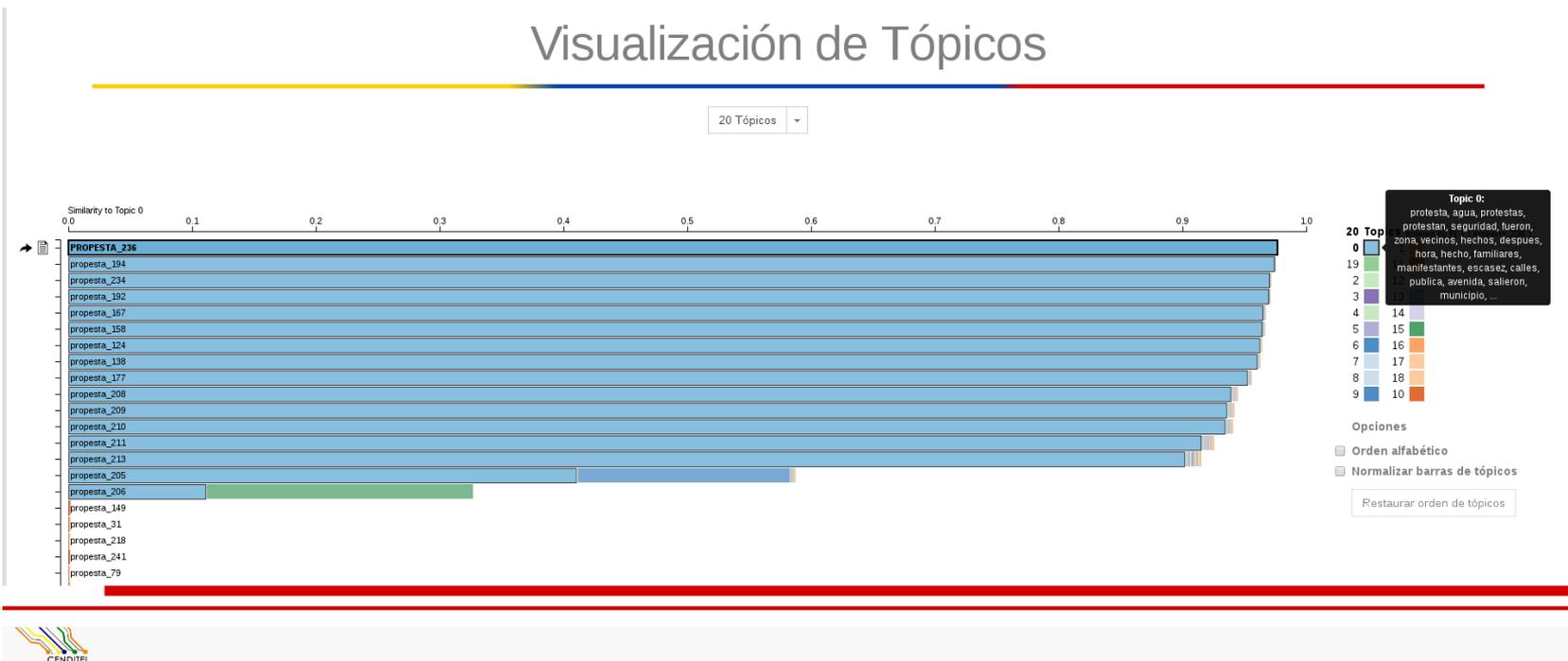
# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## MODELADO DE TÓPICOS DE UN CORPUS MEDIÁTICO

### VISUALIZACIÓN DE TÓPICOS

La herramienta permite identificar los tópicos y también los textos específicos del corpus según lo cercano que sean al tópico identificado. En este caso se identifica el tópico 0, relativo a *protestas*. Las barras permiten identificar los textos que más se acercan a este tópico, siendo el texto 236 el más cercano (el que se muestra en el borde superior del gráfico).

### Visualización de Tópicos



# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## MODELADO DE TÓPICOS DE UN CORPUS MEDIÁTICO

### VISUALIZACIÓN DE TEXTOS SEGÚN EL TÓPICO DE INTERÉS

La herramienta permite también visualizar cada texto según la cercanía con el tópico de interés e ir al texto particular para identificar los términos del mismo que conforman el tópico identificado. El tamaño de las palabras indica la frecuencia de aparición de la misma en el tópico (a mayor tamaño, mayor presencia de la palabra en el tópico), y el color permite identificar el tópico al que pertenece).

propesta\_236

minero sobreviviente confirma balacera y muertes en yacimiento de oro en tumeremo la mañana de el viernes los mineros viajaron a la de el miamo , en la frontera de los municipios roscio y sifontes en el estado bolivar , publica , en el camino , a una hora en moto , quedan los hatos san ramon , peregrino y atenas en este ultimo ocurrieron parte de los hechos .

algunos de los que iban en la via fueron interceptados en alcabalas improvisadas , supuestamente pertenecientes a la banda de , criminal conocido por su dominio en las minas de sifontes . pero otros , los mineros que iban hacia la , no corrieron con la misma suerte . esos fueron agrupados en la mina y dispersados con rafagas de tiros , como suelen hacer estos grupos hamponiles para hacer se con el control de un yacimiento . algunos , segun cuentan sobrevivientes , fueron cortados con motosierras . los que huyeron se fueron monte adentro , como el muchacho que la mañana de este lunes cuenta lo que vivió . no quiere dar su nombre . ni siquiera lo desean los familiares de los desaparecidos , " porque despues ellos hacen listas y llaman a uno : , y nos matan " . eso dice un señor mayor , canoso , casi gritando a el comandante de el destacamento de el comando rurales 629 , gregory yusti . hay mucho temor en el pueblo , pero mas aun entre

la protesta que se mantiene en la carretera nacional , la troncal 10 . el joven , que ahora anda en moto dispuesto a localizar a sus compañeros , relato que cuando llegaron los de le dio tiempo de correr . " uno no sabe ni siquiera cuantos son , lo dejaron crecer demasiado " , comenta . una de las mujeres que reclama a sus familiares , cristal cañas , dice que personas de confianza le dijeron que a los mineros los obligaron a cargar con los fallecidos y meter los en un camion azul tipo volteo . " a mi me dijeron que el que iba con mi hermano lo amarraron por los pies y lo arrastraron con una camioneta " , cuenta otra muchacha , de lentes oscuros , que duda cuando le pregunto si puede dar su nombre . en la manifestacion hay vecinos que ya han perdido el miedo . " ¡ ya basta ! ¡ ya basta de tanto miedo ! ¿ que estamos esperando ? " , reclama lisbeth quevara , quien tiene tres parientes desaparecidos . el miedo y la incertidumbre se triplican en el caso de lisbeth , pero el caso de ( seudónimo para proteger su integridad ) es de una silente felicidad : su hijo ( de quien resguarda el nombre por su seguridad ) aparecio . vivo . ileso . vivo porque sobrevivió a la balacera que confirma . vivo porque no era su día . vivo

Ver Todos Limpiar

0	1	2
19	11	3
2	12	4
3	13	5
4	14	6
5	15	7
6	16	8
7	17	9
8	18	10
9	19	20

# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## MODELADO DE TÓPICOS DE UN CORPUS DE CONSULTA PÚBLICA

### DEFINICIÓN DE CORPUS

A partir de la consulta pública abierta en torno a la propuesta de Plan de Desarrollo Nacional (Plan de la Patria) presentada en el año 2012 por el Presidente Hugo Chávez, se recogieron cerca de 10 mil propuestas populares para aportar a la construcción colectiva del Plan de Desarrollo de la Nación (2013-2019).

De este total de aportes recibidos mediante consulta abierta, se identificaron 4.634 participaciones que por su formato pudieron ser agrupadas en un corpus factible de analizar mediante el modelado de tópicos.

Los resultados obtenidos mediante la aplicación del LDA al procesamiento del corpus Plan de la Patria mostraron consistencia semántica en la identificación de los tópicos relevantes dentro del corpus, así como en la identificación de la relevancia de las propuestas dentro de cada tópico (por su grado de proximidad semántica con respecto al tópico), y la identificación de la relevancia de cada palabra dentro de los tópicos. Se obtuvieron resultados que permitieron identificar desde los 10 tópicos más relevantes hasta 90 tópicos, lo que resulta de gran utilidad en un corpus tan amplio y de naturaleza semántica tan diversa como un consulta pública relativa al Plan Nacional de Desarrollo, lo que contempla todas las áreas de competencia del Estado y las áreas de interés de las y los ciudadanos.

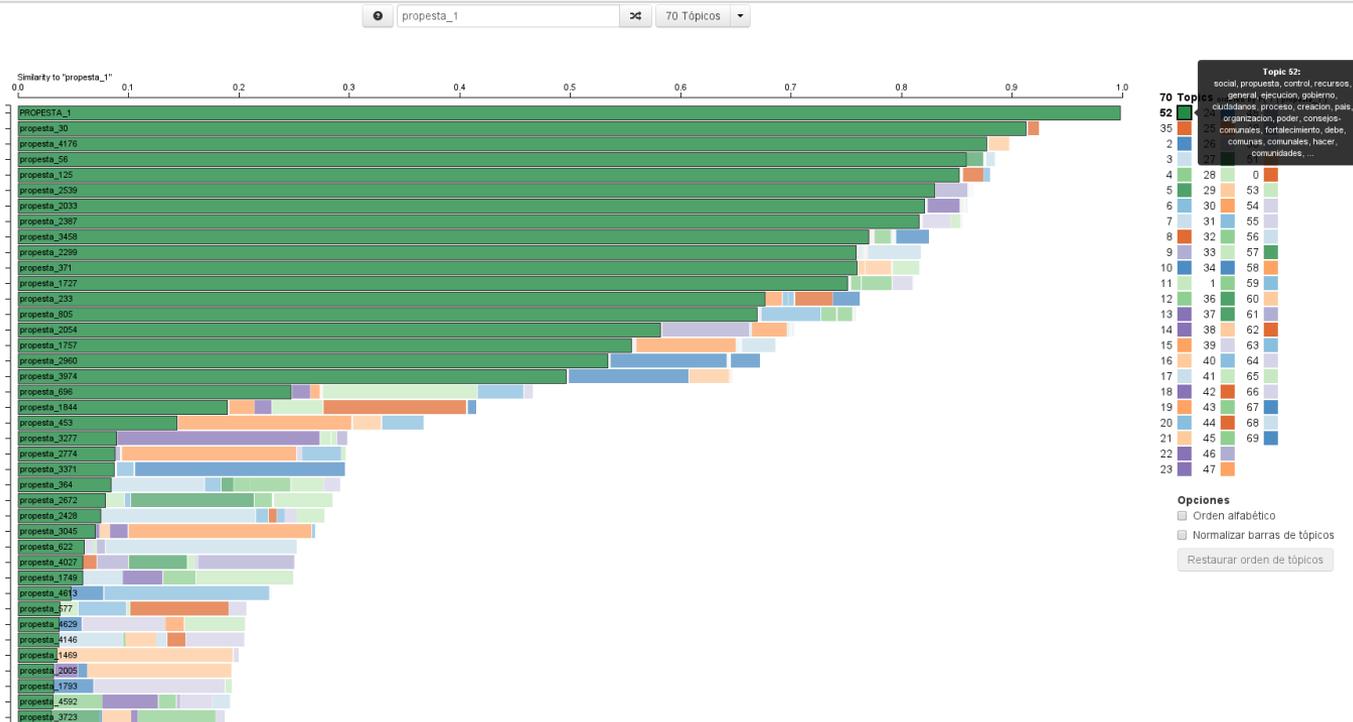
Muchos de estos tópicos permitían realizar una clasificación rápida de las propuestas según el tema que desarrollaban, lo que sería de gran ayuda para redireccionar las mismas a las instancias del Estado competentes, o para obtener datos relevantes sobre la percepción de la colectividad de los principales temas en los que debe enfocarse el Estado para el desarrollo de la Nación.

# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## MODELADO DE TÓPICOS DE UN CORPUS DE CONSULTA PÚBLICA

### VISUALIZACIÓN DE TÓPICOS

La herramienta permite visualizar desde 10 hasta 100 tópicos relevantes dentro del conjunto de documentos que conforman el corpus, así como los documentos más cercanos a los tópicos identificados. En este caso el documento 1 es el más cercano al tópico 52.



**Tópico 52**  
Relativo a **poder popular**  
(*consejos comunales, comunas, comunidades, organización, recursos, proceso, creación*)

# SISTEMA DE CONSULTA ABIERTA AUTOMATIZADA MÓDULO DE ANÁLISIS DE TEXTOS

## MODELADO DE TÓPICOS DE UN CORPUS DE CONSULTA PÚBLICA

### VISUALIZACIÓN DE TEXTOS

La herramienta permite visualizar desde 10 hasta 100 tópicos relevantes dentro del conjunto de documentos que conforman el corpus, así como los documentos más cercanos a los tópicos identificados.

propesta\_1

\*\*\* descripción : la **potencia** constituyente para volver se acto exige objetivar se en **nuevas instituciones** de **poder** y así como se hizo **necesario** la adecuación jurídica de las relaciones de los estados a partir de el mandato teórico constitucional , también es **necesario** que el **poder** constituyente reorganice sobre **nuevas** bases institucionales su propio **poder** de aquí surge la figura de los observatorios de el **poder popular** como una expresión organizada de el pueblo para el ejercicio de sus **derechos** y deberes en el **control** de la **gestión pública** y comunal , desde su propio escenario comunal , por ser uno de los **espacios** , donde se **debe** profundizar el ejercicio pleno de la democracia protagónica , para ellos se propone dos **espacios** de articulación e **integración** : un primer **nivel municipal** mediante la **conformación** de el observatorio **socialista** de el **poder popular** , como una estructura abierta , circular , dinámica , articuladora que incorpore a **voces comunales** , **habitantes entes** , **organizaciones populares** en sus prácticas de contraloría **social** sin que nadie domine o reduzca a otros , conservando su diversidad de experiencias , puntos de vista y **programas** , este observatorio tendrá como **objetivos** establecer un sistema de **información** que requieran las **organizaciones populares** de el **municipio** carlos arvelo que avanzan en la **conformación** de el **poder comunal** para el **seguimiento** y evaluación de los **programas** , **proyectos** y **políticas** que ejecuten los **entes** de el **sector público** en la localidad , en relación con su efectividad con el **proceso** de **cambio** a el **socialismo** , la autogestión de las instancias de el **poder popular** y **generar** insumos para la toma decisiones efectivas y eficientes dirigidas a la **consolidación** de los autogobiernos **populares** , nodos de contraloría **social** a **nivel** de cada **municipio** , capaz de asociar un **segundo nivel** en red de articulación que se propone es **nivel** estatal y nacional que incorpore los **diferentes** actores de el **control** de la **gestión pública** : las contralorías **sociales** , el consejo moral republicano , los tribunales de la **república** , las fiscalías , el stma nacional de **control** fiscal , la sunai , la sudcoop , las contralorías de estado y las contralorías municipales , las oficinas de auditoría , todos los órganos y **entes** de las instancias jurisdiccionales de **control** administrativa , fiscal , **penal** , **policial** , etc. , donde sus vasos comunicantes sea , con un peso específico , la **información** que genera la práctica de la contraloría **social** y donde la oficina central de **Control y seguimiento** sea el solvente que garantice la capilaridad de los vasos comunicantes que integre esta red y asuma la coordinación , mediante la **formación** de un foco que articule con los observatorios socialistas de el **poder popular** a **nivel** de cada **municipio** a el respecto , este **segundo nivel** de articulación se aproxima a la **propuesta** de el observatorio nacional de **ciencia** , **tecnología** e innovación de el mcty : " el sistema nacional de observatorios **socialista** implica el abordaje de un conjunto de temas y tareas , con el **objetivo** de **generar** una estructura sociopolítica para la captación de **información** y análisis de el efecto o **impacto** de la aplicación de **políticas públicas** ( o de la ausencia de ellas ) con el **objetivo** de orientar a el estado **venezolano** en la formulación o **ejecución** de sus **planes** , **programas** y **proyectos** , en la transición a el **socialismo** de el siglo xxi " . \*\*\* justificación : es en las localidades de el interior de el **pais** donde se exacerbó la **funcion** depredadora de el imperialismo pues **encuentra** mano de obra barata , **recursos naturales** , una **poblacion** con **trabajo** asalariado como única **forma** de subsistencia y una burocracia complice que conforma con la oligarquía **local** el **centro local** de explotación , a el cual tienen que plegar se como aprendices u operarios los **consejos comunales** y las nacientes **organizaciones populares** , en busca de **recursos** , **empleos** , aprobaciones , prebendas , etc. , con un pueblo que aun no se reconoce ni es reconocido por las **instituciones** como **parte** de el estado y corresponsable de la **vida pública** , porque de hecho no tiene mecanismos efectivos para **hacer** reales sus denuncias , ni de **derecho** para sancionar a los funcionarios **públicos** deshonestos . por el contrario existen mecanismos burocráticos que entran y demoran la **participacion** ciudadana ejemplo , el silencio administrativo de las **zonas** educativas de el ministerio de **educacion** , de fundacomunal carabobo , taquilla única , min **comunales** , redes , pae , contraloría **municipal** , quienes ignoran o desconocen los **recursos** jerárquicos de la topa , la ley organica de transferencia de competencias a el **poder popular** , la ley contra la corrupción , la ley de contrataciones , etc. porque no proceden las denuncias en fiscalía de el **poder popular** por ? porque hay dilación en la jurisdicción contenciosa administrativa ? donde **están** los fiscales especiales que dice la locc fueran nombrados para atender la violación de la **participacion** protagónica ? porque los **voces** municipales son designados por consejo federal de **gobierno** y no el **poder popular** ? como agravante a el quiebre moral de las **instituciones** , hay que agregar que las **nuevas formas** de **organizacion** popular nacieron burocráticas que lejos de **promover** su autonomía , apuestan a su fracaso abiertamente y en otros **casos** , por su burocratismo , entrambamiento , falta de transparencia e incluso de **formacion** técnica , legal o **politica** , aceleran el fracaso de las **nuevas formas** de **organizacion** popular , una **situacion** diferente hubiera sido que las **nuevas organizaciones populares** no nacieran , sino con el de el estado , mas identificadas con el proyecto **politico** ( **caso** )

### Visualizador de textos

El visualizador permite identificar rápidamente por el tamaño y color de las palabras los tópicos que conforman el documento.

Ver Todos Limpiar

