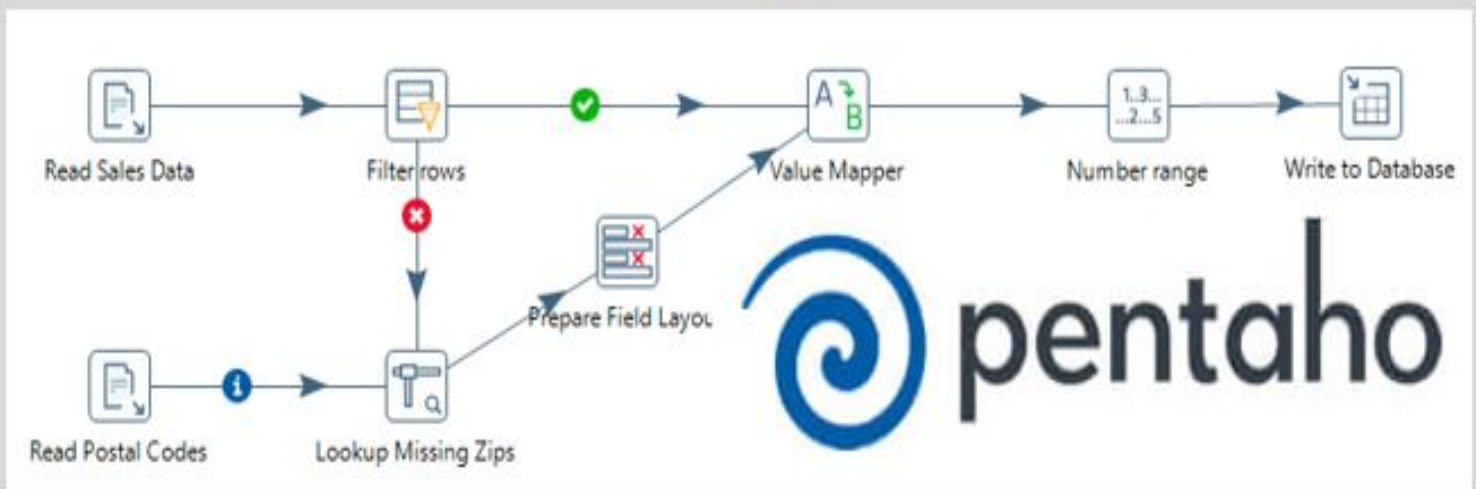
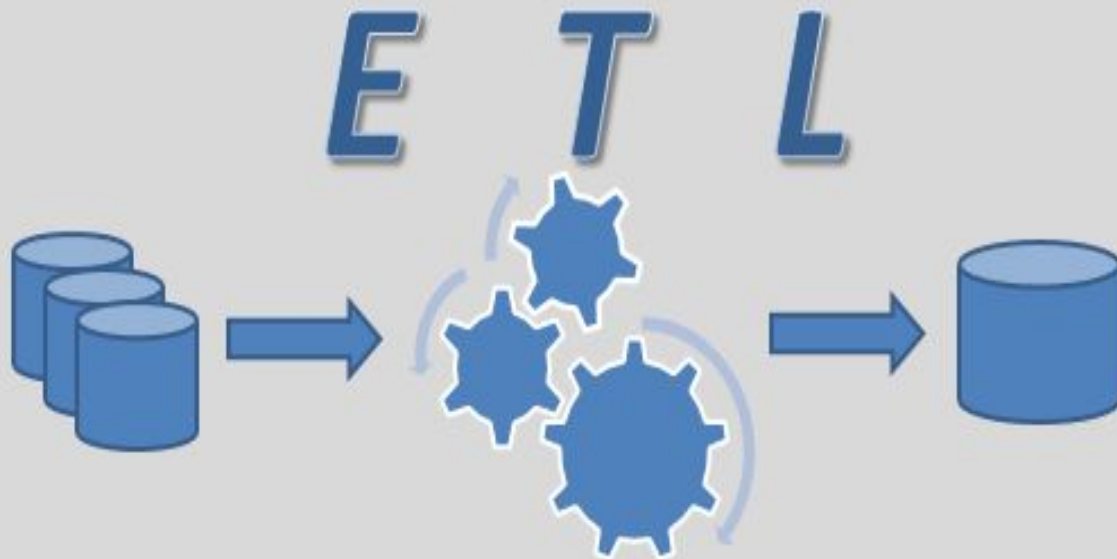


PROYECTO DE COMPUTACION I

PENTAHO

ACTIVIDAD 2



Índice

Contextualización	3
Descripción de las fuentes de datos	4
Descripción del proceso	7
Resultados	14
Conclusión	16
Bibliografía	17

Contextualización

El proyecto que realizaremos se base en torno a la idea de la España vaciada, en el cual intentaremos frenar y solucionar los problemas sociales y económicos causados por este fenómeno, principalmente causados por la desigualdad en cuanto al reparto de la población española en donde los pueblos más alejados de los núcleos urbanos acaban teniendo mayoritariamente población envejecida y acaban desapareciendo.

Es por ello que utilizaremos un proceso ETL en donde moveremos información de diversas fuentes como medios de noticias, transformandolas para que sea entendible al usuario y cargando todo en nuestro data lake, ayudando así al usuario que quiera establecerse en la zona o uno de estos pueblos y ofrecerle toda la información necesaria para que no acabe en una situación desfavorable en la que le sea muy complicado realizar las tareas del día a día..

Descripción de las fuentes de datos

Para el proyecto he utilizado medios de noticias influyentes en España, de las que se encuentran los conocidos: El País, Cope, Antena 3 y La Sexta. He escogido estos medios porque al ser de los más grandes de España podemos considerar que la información es fiable puesto que todos sus datos son sacados de fuentes fiables como puede ser el INE (Instituto Nacional de Estadística), además contienen más información respecto al tema de la España vaciada que medios de noticias más pequeños, enriqueciendo nuestra página con más información relevante.

Al utilizar sus páginas web, todas son páginas HTML y por lo tanto siguen la estructura DOM por lo que aprovechándonos de eso, para la gran mayoría de información que requerimos se encuentra en el <head>, en la forma de <meta> tags que se utilizan para codificar información que luego es interpretada por el navegador y los buscadores para clasificar su contenido. Para la estrategia que he planteado, todo menos el cuerpo de la noticia podría sacarse del header, pero he aprovechado el uso de los JSON-LD en estas páginas para ello aunque podría haber sacado toda la información de ahí.

JSON-LD (JavaScript Object Notation for Linked Data) es un método basado en JSON para añadir datos estructurados a las páginas web en donde los metadatos se implementan dentro de una etiqueta <script> separada del contenido bien en el <head> o en el <body> del documento HTML. Como los motores de búsqueda necesitan metadatos para ayudar a procesar la información, utilizamos el JSON-LD para extraer la información requerida de forma cómoda y sencilla mediante el uso de simples regex. Centrándonos en los datos que no constituyen el cuerpo, las keywords y la fecha de publicación:

- En **Antena 3** para obtener el título usamos el tag `<meta property="og:title" content="¿Ayudaría a la España vaciada un Ministerio para la despoblación?">`, para la entrada el tag `<meta property="og:description" content="Los vecinos de la España vaciada han explicado qué le solicitarían al Ministerio para la despoblación si existiese.">` y para la url el tag `<meta`

```
property="og:url"
content="https://www.antena3.com/noticias/sociedad/ayudaria-a-l
a-espana-vaciada-un-ministerio-para-la-despoblacion\_201911165dd
03c3c0cf2c8d2dbb2be3f.html">
```

- En **Cope** para obtener el título usamos el tag `<meta property="og:title"`
`content="Cinco iniciativas que buscan repoblar la España`
`vaciada" />` , para la entrada el tag `<meta property="og:description"`
`content="La vivienda, el reparto a domicilio o las nuevas`
`tecnologías son algunos de los ámbitos que más soluciones`
`arrojan" />` y para la url el tag `<meta property="og:url"`
`content="`
- En **La Sexta** para obtener el título usamos el tag `<meta property="og:title"`
`content="Paro, despoblación y un futuro incierto: León empieza`
`a sufrir las consecuencias de la España vaciada"/>`, para la
entrada el tag `<meta property="og:description"` `content="León ya es`
`definida como la 'Detroit española'. El paro roza el`
`17% en la capital y el 19% en Ponferrada, la segunda ciudad más`
`poblada de la provincia. "/>` y para la url el tag `<meta`
`property="og:url"`
`content=".`

Por otro lado en cuanto al cuerpo, las keywords y la fecha de la noticia aprovechandonos del formato JSON (key-value/clave-valor pair/pareja) del JSON-LD buscamos la pareja clave-valor para ellos de la siguiente manera:

- En **Antena 3** para el cuerpo usamos la pareja clave-valor `"articleBody": "Los`
`vecinos de la España vaciada reclaman..."`, para la fecha usamos la
pareja `"datePublished": "2019-11-16T19:13:16.219+01:00"` y para las

keywords la pareja "keywords": ["España despoblada", "España vaciada", "video"].

- En **Cope** para el cuerpo usamos la pareja clave-valor "articleBody": "En tiempos en los que la masificación...", para la fecha usamos la pareja "datePublished": "2020-03-28T02:10:00+01:00" y para las keywords la pareja "keywords": ["pueblos", "Actualidad", "La españa vaciada"].
- En **La Sexta** para el cuerpo usamos la pareja clave-valor "articleBody": "El abandono de la agricultura y la minería...", para la fecha usamos la pareja "datePublished": "2020-06-20T22:30:00.270Z" y para las keywords la pareja "keywords": ["España", "Despoblación", "España vacía"].

Descripción del proceso

El proceso para realizar las transformaciones necesarias para obtener los ficheros de textos con la información requerida delimitada por sus campos respectivos ha seguido la misma estrategia general, así que empezando por el proceso para las noticias de La Sexta:

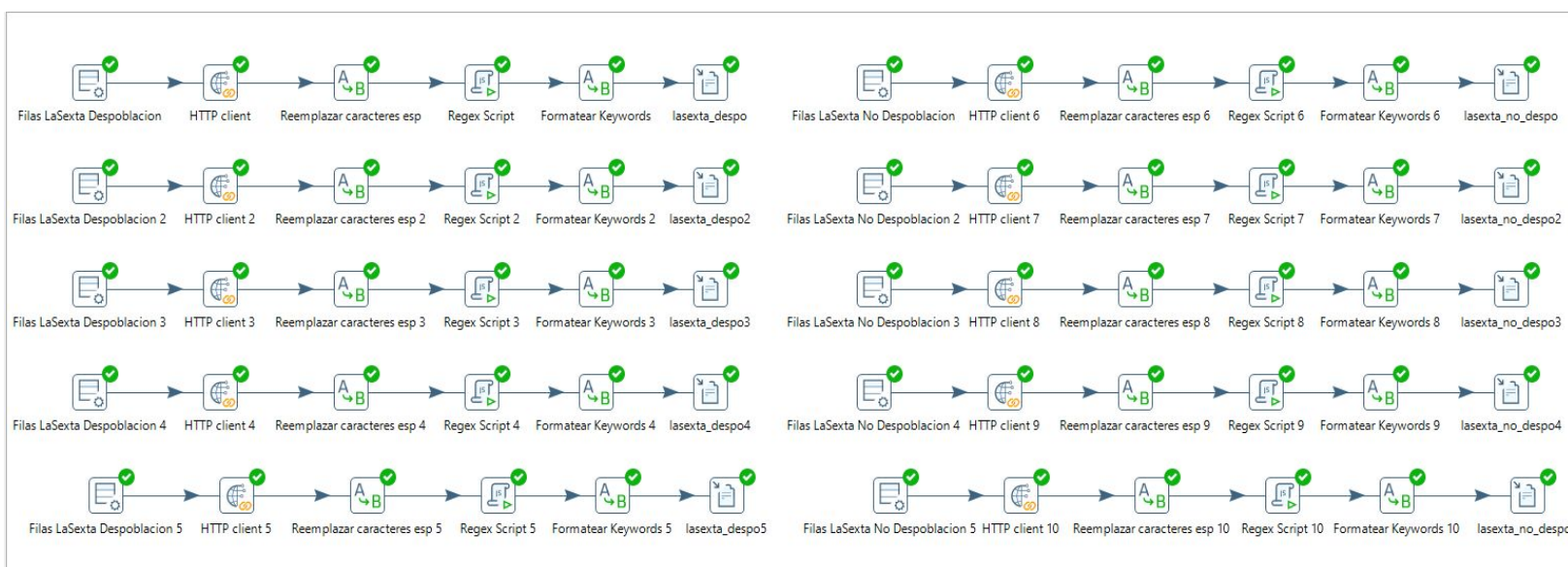


Figura 1

Transformación de noticias La Sexta

Como se puede observar en la figura 1, el esquema general con los pipes para cada tarea es igual para cada noticia, donde vemos 5 para aquellas relacionadas con la despoblación y 5 para aquellas relacionadas con la no despoblación. Empezando por conseguir la página HTML en formato txt dividiéndola en filas, aplicando un reemplazamiento de strings para reemplazar caracteres unicode del estilo `#&39;` con su respectiva `'`, aplicando las regex necesarias y volviendo a reemplazar ciertos parámetros para una salida a fichero de texto más elegante y cómoda de leer desde el punto de vista del usuario.

Empezando por el paso de generar filas, le pasamos la URL para que divida el código HTML en filas además del nombre del paso para poder utilizar su resultado como salida más adelante, como se puede ver en la figura 2:

Generate rows

Step name: Filas LaSexta Despoblacion

Limit: 1

Never stop generating rows: ☐

Interval in ms (delay): 5000

Current row time field name: now

Previous row time field name: FiveSecondsAgo

Fields:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value
1	LaSexta	String							https://www.lasexta.com/noticias/socieda

Help OK Preview Cancel

Figura 2

Paso para generar filas

Para el siguiente paso, como se puede apreciar en la figura 3 obtenemos la URL de la salida del paso previo para conectarse a la página y poder descargar el código ya en formato de texto como nos interesa, aplicando un nombre a la salida para poder usarla en el siguiente paso:

HTTP client

Step name: HTTP client

General Fields

Settings

URL:

Accept URL from field? ☒

URL field name: LaSexta

Encoding (empty means standard): UTF-8

Connection timeout: 10000

Socket timeout: 10000

Connection close wait time: -1

Output fields

Result field name: http_result

Figura 3

Paso para obtener el código HTML

Como estamos hablando de HTML hay mucho uso de caracteres unicode, y no están en el formato que es legible por una persona, por lo que en este paso remediamos esto cambiándolo a un formato más comprensivo:

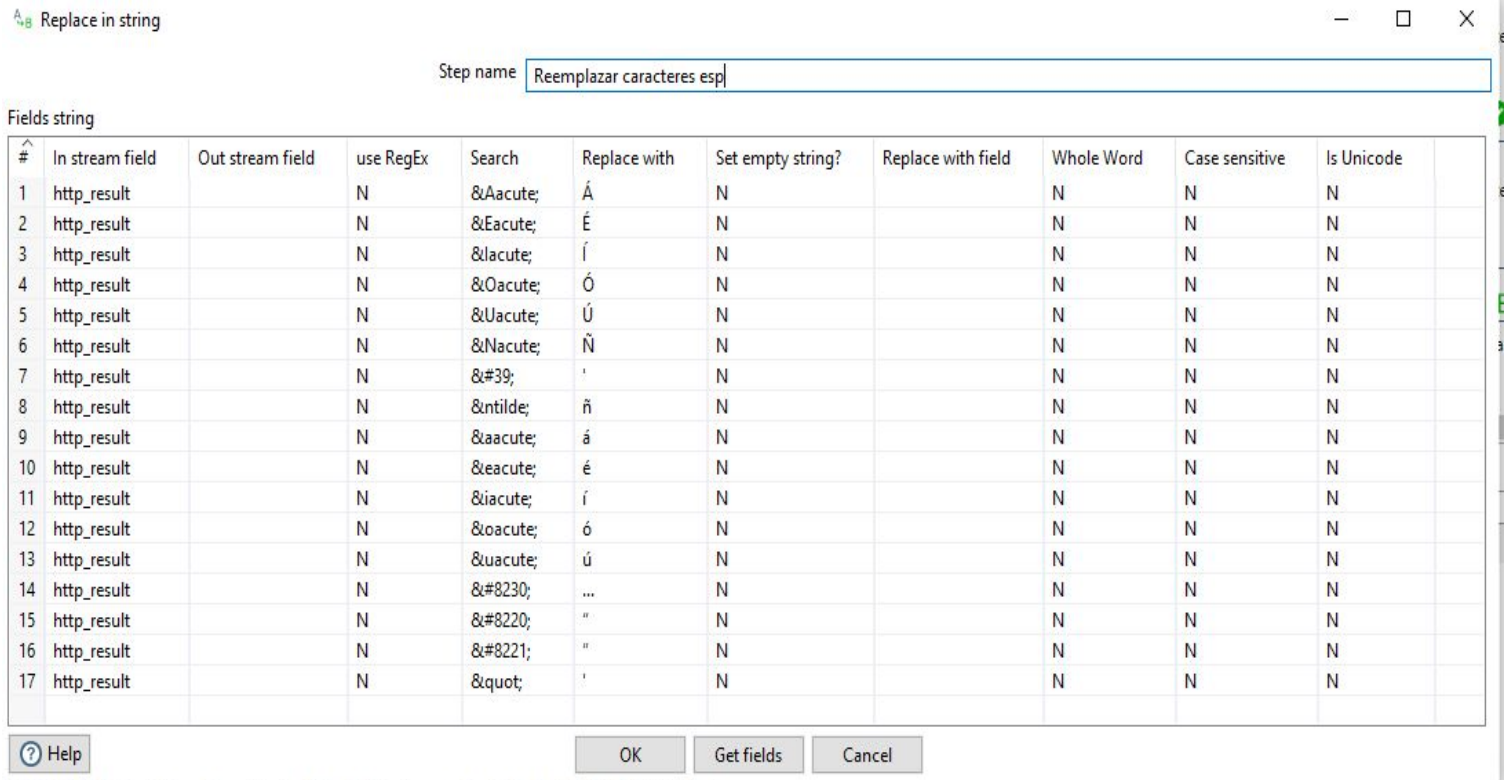
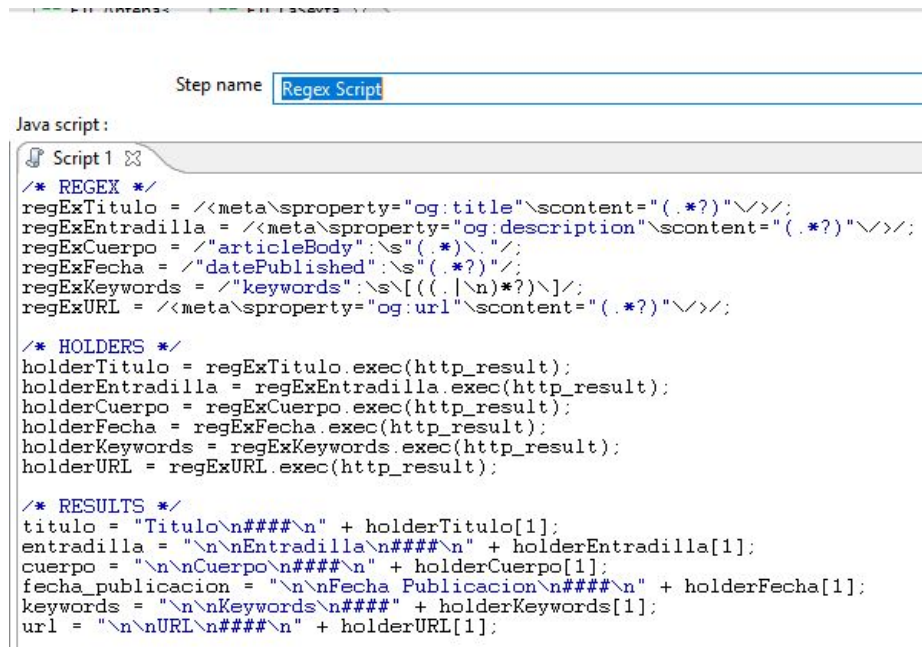


Figura 4
Cambiando el formato unicode

En este punto ya tenemos el código HTML en formato texto habiendo quitado todos los caracteres unicode que no estaban en el formato que deseábamos, por lo que ahora nos queda la parte en la que obtenemos la información que queremos mediante el uso de RegEx, ya que sabiendo que sigue la estructura DOM y toda la información está comprimida en tags.

En cuanto a las RegEx utilizadas son todas iguales para todas las transformaciones excepto por detalles mínimos como puede ser un espacio o un punto al final o al principio de los atributos de los tags, pero este se debe a la forma de programación de cada página, aunque esto solo ocurre con Cope ya que con La Sexta y Antena 3 se puede aplicar las mismas RegEx sin ninguna problema.



```
Step name: Regex Script

Java script:

Script 1
/* REGEX */
regExTitulo = /<meta\sproperty="og:title"\scontent="(.*?)"/>/;
regExEntradilla = /<meta\sproperty="og:description"\scontent="(.*?)"/>/;
regExCuerpo = /"articleBody":\s"(.*?)"/;
regExFecha = /"datePublished":\s"(.*?)"/;
regExKeywords = /"keywords":\s"([(|\n|)*)"/;
regExURL = /<meta\sproperty="og:url"\scontent="(.*?)"/>/;

/* HOLDERS */
holderTitulo = regExTitulo.exec(http_result);
holderEntradilla = regExEntradilla.exec(http_result);
holderCuerpo = regExCuerpo.exec(http_result);
holderFecha = regExFecha.exec(http_result);
holderKeywords = regExKeywords.exec(http_result);
holderURL = regExURL.exec(http_result);

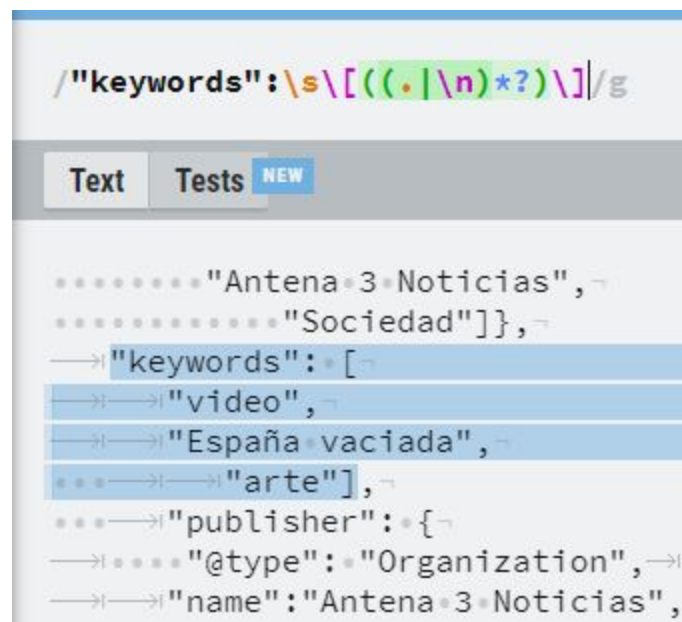
/* RESULTS */
titulo = "Titulo\n####\n" + holderTitulo[1];
entradilla = "\n\nEntradilla\n####\n" + holderEntradilla[1];
cuerpo = "\n\nCuerpo\n####\n" + holderCuerpo[1];
fecha_publicacion = "\n\nFecha Publicacion\n####\n" + holderFecha[1];
keywords = "\n\nKeywords\n####\n" + holderKeywords[1];
url = "\n\nURL\n####\n" + holderURL[1];
```

Figura 5

Expresiones RegEx

En cuanto a la RegEx para el título aprovecha que el tag para el título siempre es el mismo, por lo que podemos usar el formato `<meta property="og:title" content=""/>`; en donde solo debemos especificar que entre las comillas del content puede ir cualquier cosa, pero nótese cómo utilizamos un lazy search añadiendo un símbolo de interrogación a el asterisco (.*?) para que no busque hasta la última ocurrencia del “ si no que pare en cuanto encuentre el primero, ya que si no obtendremos mucho más que el título, y esto es igual para el resto de RegEx excepto el articleBody que no lo necesita. La segunda RegEx para la entrada también utiliza la misma técnica, en donde siempre va a ser `meta property="og:description" content=""/>`; donde podemos obtener cualquier cosa dentro de los asteriscos al igual que la tercera (cuerpo), la cuarta (fecha) y la última (url) ya que siempre va a haber un patron así.

Por otro lado en RegEx de las keywords ya que el formato se encuentre en JSON, por lo que tenemos que aplicar una regex que aplica a multi-lineas dentro de un array, por lo que lo que decimos es que al igual que las anteriores sigue un patrón fijo del estilo: "keywords: []" con varios saltos de línea por cada elemento del array, y para eso establecemos: `((.|\n)*?)` en donde decimos que obtengamos cualquier cosa o un salto de línea, 0 o más veces por el `*`, que también aplicamos un lazy search para que no nos pasemos hasta el último corchete de otra cosa y obtengamos información de más, como se puede ver en la figura posterior:



The screenshot shows a web-based regex testing interface. At the top, a text input field contains the regex pattern: `/"keywords":\s\[((.|\n)*?)\]/g`. Below the input, there are two tabs: "Text" and "Tests" (with a "NEW" button). The "Text" tab is active, displaying a multi-line JSON string. The string is:

```
....."Antena 3 Noticias",
....."Sociedad"]}],
->"keywords": [
->->"video",
->->"España vaciada",
*->->"arte"],
*->"publisher": {
->->->"@type": "Organization",
->->->"name": "Antena 3 Noticias",
```

 The regex pattern is designed to match the "keywords" array in this JSON structure, capturing all elements including newlines.

Figura 6
Ejemplo RegEx array multi-línea

Finalmente como tenemos varios tabs y espacios no muy cómodos de leer cuando los escribas al fichero de texto aplicamos una regex final para reemplazarlos por nada, en donde aplicamos la regex: **(?!\\n)(\\t+|\\s{2,})** donde buscamos dos grupos, el primero es un negative lookahead donde decimos que sí coincide con un salto de línea no lo incluya en el resultado, mientras que en el segundo grupo quitamos todos los tabulados y más de dos espacios juntos (ya que no queremos quitar el espacio entre las palabras) para dejarlo en un formato más legible por una persona, como se puede ver en la figura 6:

Replace in string

Step name: Formatear Keywords

Fields string

#	In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with field	Whole Word	Case sensitive	Is Un
1	keywords	keywords_formateado	Y	(?!\\n)(\\t+ \\s{2,})		N		N	N	N

Figura 7

RegEx final

Visto mejor en la figura posterior:



Figura 8

Ejemplo RegEx negative look-ahead

Como el resto de transformaciones siguen exactamente la misma estrategia y mismas regex (a pesar de mínimas modificaciones como un espacio o un punto como he mencionado anteriormente) no las explicare en detalle, pero cabe mencionar que en el caso de el medio de noticias Cope al tener la información más “comprimada” el array de keywords está en la misma línea, simplificando aún más le RegEx y quitando el paso de formatearla luego:

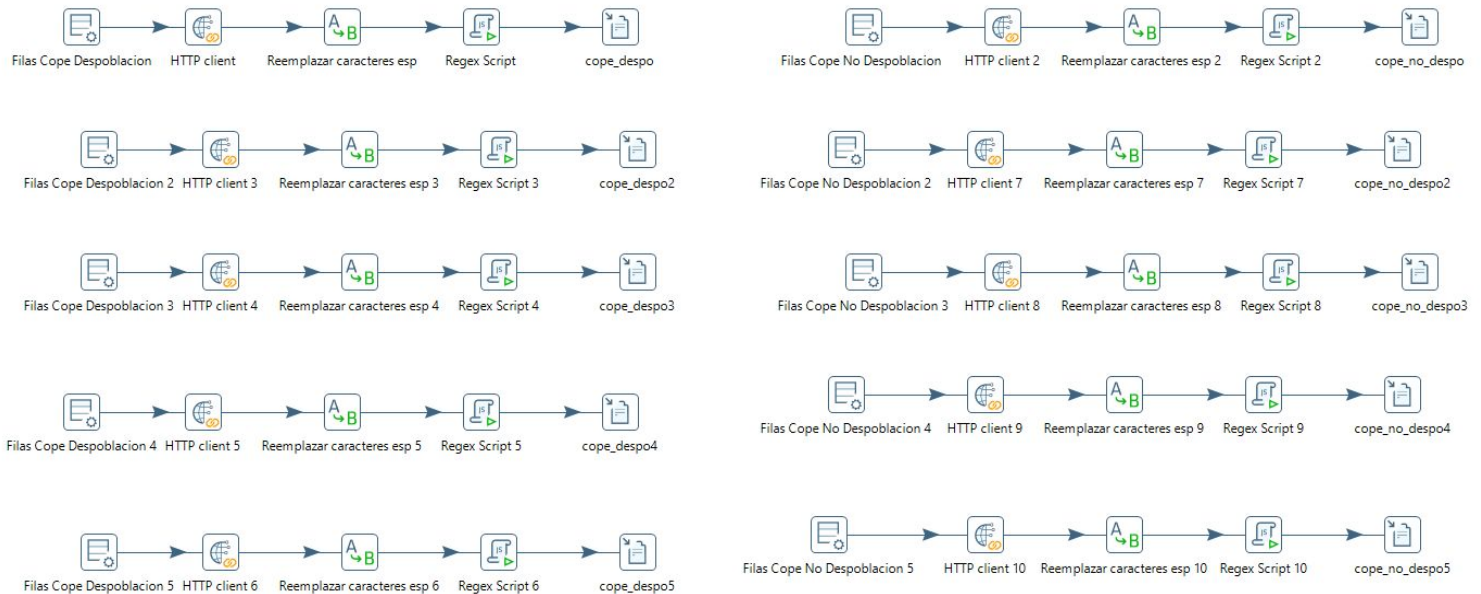


Figura 9

Transformacion noticias Cope

La regex mencionada anteriormente es del estilo: **"keywords":\s\[(".*?")\]** donde vuelve a aplicar el lazy search y obtiene cualquier carácter que esté entre los corchetes.

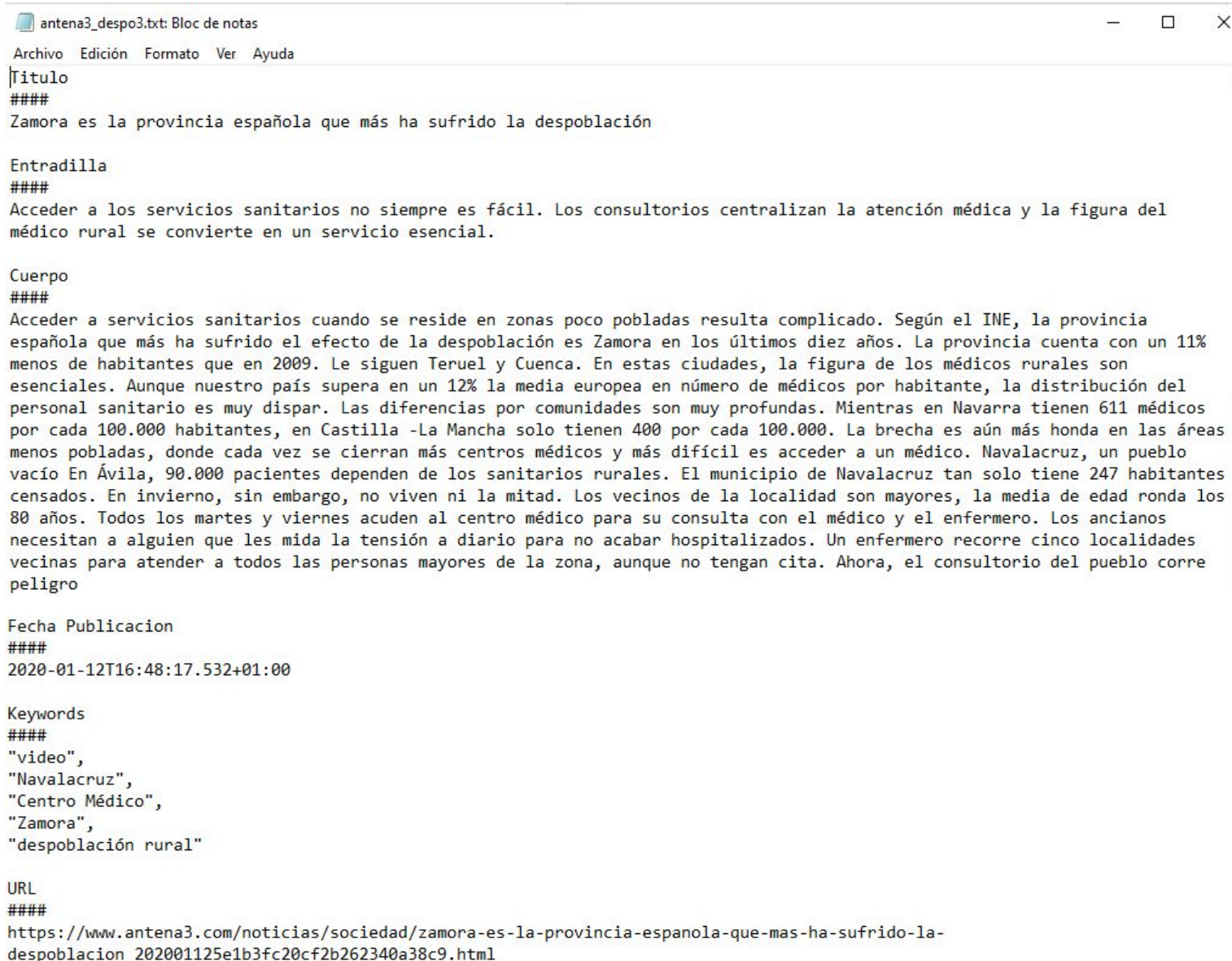
```
/* REGEX */
regExTitulo = /<meta\sproperty="og:title"\scontent="(.*?)"\s\/>/;
regExEntradilla = /<meta\sproperty="og:description"\scontent="(.*?)"\s\/>/;
regExCuerpo = /"articleBody":\s"(.*?)"\s\/;
regExFecha = /"datePublished":\s"(.*?)"\s\/;
regExKeywords = /"keywords":\s\[(".*?")\]/;
regExURL = /<meta\sproperty="og:url"\scontent="(.*?)"\s\/>/;
```

Figura 10

RegEx noticias Cope

Resultados

Los resultados obtenidos son 30 ficheros en formato .txt donde hay 5 de despoblación y otros 5 de no despoblación por cada medio de noticia (Cope, Antena 3 y La Sexta) en donde la categoría de los artículos de no despoblación varían desde deportes y actualidad hasta tecnología. Un ejemplo de una noticia de despoblación visto en la figura posterior:



```
antena3_despo3.txt: Bloc de notas
Archivo Edición Formato Ver Ayuda
Titulo
####
Zamora es la provincia española que más ha sufrido la despoblación

Entradilla
####
Acceder a los servicios sanitarios no siempre es fácil. Los consultorios centralizan la atención médica y la figura del médico rural se convierte en un servicio esencial.

Cuerpo
####
Acceder a servicios sanitarios cuando se reside en zonas poco pobladas resulta complicado. Según el INE, la provincia española que más ha sufrido el efecto de la despoblación es Zamora en los últimos diez años. La provincia cuenta con un 11% menos de habitantes que en 2009. Le siguen Teruel y Cuenca. En estas ciudades, la figura de los médicos rurales son esenciales. Aunque nuestro país supera en un 12% la media europea en número de médicos por habitante, la distribución del personal sanitario es muy dispar. Las diferencias por comunidades son muy profundas. Mientras en Navarra tienen 611 médicos por cada 100.000 habitantes, en Castilla -La Mancha solo tienen 400 por cada 100.000. La brecha es aún más honda en las áreas menos pobladas, donde cada vez se cierran más centros médicos y más difícil es acceder a un médico. Navalacruz, un pueblo vacío En Ávila, 90.000 pacientes dependen de los sanitarios rurales. El municipio de Navalacruz tan solo tiene 247 habitantes censados. En invierno, sin embargo, no viven ni la mitad. Los vecinos de la localidad son mayores, la media de edad ronda los 80 años. Todos los martes y viernes acuden al centro médico para su consulta con el médico y el enfermero. Los ancianos necesitan a alguien que les mida la tensión a diario para no acabar hospitalizados. Un enfermero recorre cinco localidades vecinas para atender a todos las personas mayores de la zona, aunque no tengan cita. Ahora, el consultorio del pueblo corre peligro

Fecha Publicacion
####
2020-01-12T16:48:17.532+01:00

Keywords
####
"video",
"Navalacruz",
"Centro Médico",
"Zamora",
"despoblación rural"

URL
####
https://www.antena3.com/noticias/sociedad/zamora-es-la-provincia-espanola-que-mas-ha-sufrido-la-despoblacion_202001125e1b3fc20cf2b262340a38c9.html
```

Figura 11

Ejemplo noticia España vacía

Mientras que un ejemplo de un artículo hablando de la actualidad quedaría de la siguiente forma:

```
lasexta_no_despo3.txt: Bloc de notas
Archivo Edición Formato Ver Ayuda
Titulo
####
Explota una bomba casera cerca de la casa del expresidente argentino Mauricio Macri

Entradilla
####
Una bomba casera ha explotado este miércoles cerca de la casa del expresidente argentino Mauricio Macri (2015-2019), y aunque no ha dejado heridos ni daños materiales, se investiga si está relacionado con algún tipo de intimidación política.

Cuerpo
####
En la localidad bonaerense de Acassuso, a apenas 300 metros de la casa de Macri, un artefacto ha sido detonado dentro de un contenedor, que se prendió fuego tras la explosión. Hasta el momento, no se han reportado heridos por el incidente ni daños materiales más allá del propio contenedor, mientras que la Policía Federal se encuentra en la zona para determinar lo sucedido y si está relacionado con algún tipo de intimidación contra el expresidente. Según las primeras investigaciones recogidas por la prensa argentina, aunque el artefacto no tenía la capacidad de causar grandes estragos, se descarta en un principio que se tratara de una explosión accidental, sino que habría sido provocada, por lo que se investigan posibles motivaciones políticas. Vecinos de la zona han asegurado al periódico argentino 'La Nación' que la detonación ha sido 'fuerte', mientras que se ha confirmado que al momento de la explosión el antiguo mandatario se encontraba en su residencia. En este momento, las autoridades trabajan en revisar las cámaras de seguridad del entorno para identificar a posibles sospechosos, después de que varios testigos hayan asegurado que dos personas con moto se habrían acercado hasta el contenedor

Fecha Publicacion
####
2020-11-12T08:44:23.679+01:00

Keywords
####
"Mauricio Macri",
"Bomba",
"atentado",
"Noticias Última Hora",
"Argentina"

URL
####
https://www.lasexta.com/noticias/internacional/explota-una-bomba-casera-cerca-de-la-casa-del-expresidente-argentino-mauricio-macri_202011125face7d8cdfdea00017e6879.html
```

Figura 12
Ejemplo noticia actualidad

Conclusión

En vista a lo reflejado anteriormente queda claro que esta estrategia funciona y es utilizable a una escala con pocas noticias, pero si queremos ampliar este número para poder recibir más información queda claro que tenemos que automatizar de alguna manera el obtener noticias de un medio, además de empezar a buscar ciertas keywords como “españa vacía” para de verdad utilizar aquellas que nos son útil además de tener que modificar ciertas regex para poder capturar más grupos ya que no todas siguen exactamente la misma rigurosidad a la hora de nombrar los atributos y contenidos (algunas usan un espacio al final, otras un punto otras no etc.) y empezar a guardar las noticias en contenidos como un data lake.

De todos modos nos da una idea de lo que esperar a la hora de diseccionar el código HTML para buscar la información requerida y una aproximación general de cómo hacerlo, ya que de ello depende que sea una tarea relativamente fácil y mantenible o algo mucho más complicado que sea inviable en el futuro.

Bibliografía

- [1] Diego Maradona: El médico Leopoldo Luque asegura que "el alta es inminente, va a ser pronto", 2020. *Antena 3 Noticias* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.antena3.com/noticias/deportes/futbol/leopoldo-luque-medico-de-maradona-el-alta-es-inminente-va-a-ser-pronto_202011115fabb1e69a1c4e000179dcd9.html
- [2] Google Fotos: Adiós al almacenamiento ilimitado a partir de junio de 2021, 2020. *Antena 3 Noticias* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.antena3.com/noticias/tecnologia/google-fotos-obligara-pagar-almacenamiento-archivo-s-partir-junio-2021_202011125fad1350cdfdea00017eb788.html
- [3] Xbox Series X: Así es la nueva consola de Microsoft ya disponible en España, 2020. *Antena 3 Noticias* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.antena3.com/noticias/tecnologia/asi-xbox-series-nueva-consola-microsoft-disponible-espana_202011115fabbd3b2416a60001a74e0d.html
- [4] Apple: La compañía presenta su chip M1 en sociedad como "el procesador más potente jamás creado", 2020. *Antena 3 Noticias* [online], https://www.antena3.com/noticias/tecnologia/procesador-mas-potente-jamas-creado-apple-presenta-chip-sociedad_202011105faae96a0e37c40001fecaac.html
- [5] Día del Soltero 2020: Estos son los cupones descuento de Aliexpress para el 11.11, 2020. *Antena 3 Noticias* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.antena3.com/noticias/tecnologia/dia-soltero-2020-estos-son-cupones-descuento-aliexpress-1111_202011115fabe0579a1c4e00017a3b57.html
- [6] EFE, AGENCIA, 2020, Arrimadas dice que negociará hasta el final los presupuestos y pide no ceder al chantaje separatista. *COPE* [online]. [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.cope.es/actualidad/espana/noticias/arrimadas-dice-negociara-hasta-final-los-presupuestos-pide-ceder-chantaje-separatista-20201112_992955
- [7] Cardassar y Marchamalo se clasifican en penaltis a la siguiente ronda de la Copa del Rey, 2020. *COPE* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.cope.es/deportes/futbol/noticias/cardassar-marchamalo-clasifican-penaltis-siguiente-ronda-copa-del-rey-20201111_992556

-
- [8] El Bayern recupera el liderato, tras vencer el Borussia Dortmund; victorias de Wolfsburg y Leverkusen, 2020. *COPE* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.cope.es/deportes/futbol/futbol-internacional/noticias/jornada-bundesliga-2020-2021-06-11-2020-20201106_985636
- [9] El Gobierno baja a 0,72 euros el precio máximo de las mascarillas quirúrgicas, 2020. *COPE* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.cope.es/actualidad/sociedad/noticias/gobierno-baja-072-euros-precio-maximo-las-mascarillas-quirurgicas-20201112_994160
- [10] DIGITAL, REDACCIÓN, 2020, ¿Podrás viajar el fin de semana? Estas son las restricciones de movilidad que siguen vigentes. *COPE* [online]. [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.cope.es/actualidad/espana/noticias/podras-viajar-fin-semana-estas-son-las-restricciones-movilidad-que-siguen-vigentes-20201112_994126
- [11] Primark da marcha atrás y retira el ERTE para sus 7.000 empleados en toda España. *LaSexta* [online]. [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.lasexta.com/noticias/economia/primark-marcha-atras-retira-erte-sus-7000-empleados-toda-espana_202011115fac1aa4cdfdea00017d5408.html
- [12] Maricopa, el condado clave que ha dado la victoria a Biden en el feudo republicano de Arizona, 2020. *LaSexta* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.lasexta.com/noticias/internacional/elecciones-eeuu/maricopa-el-condado-clave-que-ha-dado-la-victoria-a-biden-en-el-estado-de-arizona_202011125fad2e38f1b0790001a085e5.html
- [13] Explota una bomba casera cerca de la casa del expresidente argentino Mauricio Macri, 2020. *LaSexta* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.lasexta.com/noticias/internacional/explota-una-bomba-casera-cerca-de-la-casa-del-expresidente-argentino-mauricio-macri_202011125face7d8cdfdea00017e6879.html
- [14] La solución al transporte a -80°C de la vacuna de Pfizer: así llegará a España, 2020. *LaSexta* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.lasexta.com/noticias/internacional/la-solucion-al-transporte-a-80c-de-la-vacuna-de-pfizer-asi-llegara-a-espana_202011115fac0918cdfdea00017d36ec.html

[15] Protestas en Brasil por la humillación a una víctima de violación: la defensa usó fotos de sus redes como 'prueba', 2020. *LaSexta* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: https://www.lasexta.com/noticias/internacional/protestas-en-brasil-por-la-humillacion-a-una-victima-de-violacion-la-defensa-uso-fotos-de-sus-redes-como-prueba_202011095fa972ea9a1c4e000176e896.html

[16] Marcado con JSON-LD según schema.org, 2020. *IONOS Digitalguide* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: <https://www.ionos.es/digitalguide/paginas-web/creacion-de-paginas-web/tutorial-marcado-con-json-ld-segun-schemaorg/>

[17] RegExr: Learn, Build, & Test RegEx, 2020. *RegExr* [online], [Fecha de consulta: 10 Noviembre 2020] Disponible en: <https://regexr.com/>