**TECH SPEC TFCA**

**Type Boletin: No captcha, No Security.**

1. This is the base dictionary. Before going through any filter, each type of 'document' has to be adapted to these fields, so we can put it in our ‘raw’ db:

{

"actor" :,

"demandado" :,

"entidad" :,

"expediente" :,

"fecha" :,

"fuero" :,

"juzgado" :,

"tipo" :,

"acuerdos" :,

“monto”:,

“fecha\_presentacion”:

“actos\_reclamados”:

“actos\_reclamados\_especificos”:

“naturaleza\_procedimiento”

“prestacion\_demandada”

“organo\_jurisdiccional\_origen”

“expediente\_origen”

“materia”:

“submateria”:

“fecha\_sentencia”:

“sentido\_sentencia”:

“resoluciones”:

“origen:  
“fecha insercion:

To do the insert in the database "raw". We need the data only in upper () and without any special characters ó, ö, etc. consider the following:

0.- Put in utf-8

1.- Transform 'áéíóöúüÁÉÍÓÖÚÜ' to 'aeioouuAEIOOUU'

2- Quit double space, triple space, n space, only keep space.

3.- All in upper.

1. The Scraper must have a “memory” function so as not to repeat the extraction of data from the same date.
2. The Scraper must check that there are no duplicates considering the following data:

- actor

- demandado

- juzgado

- expediente

-materia

-fecha

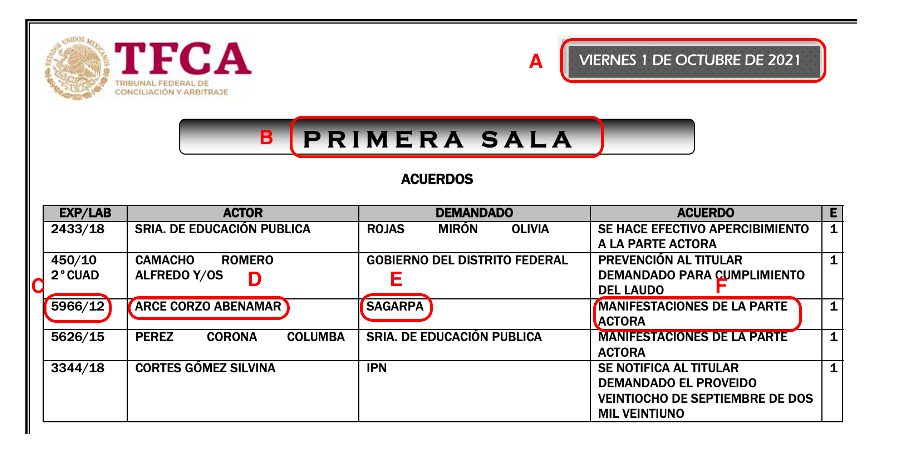
1. The code must be prepared to insert the data obtained and assigned in the Raw dictionary directly as a dict in MongoDB
2. In case of page saturation, the code must detect it and put it in timesleep for two hours and resume after this time.
3. An index must be generated with the date, time and amount of data extracted to keep a record and resume where the connection is lost.
4. When the scraper obtains all the existing data to the current date, it will have to wait 48 hours to extract new data again.
5. The scraper must have variables that control the start-date and end-date of the pages to scrap. And if the end-date variable is empty / undefined, you have to go ahead and proceed assuming end-date is today and scrap until today and then day by day to keep extracting new available data.
6. Once the scraper is working, you must send samples in json format with a minimum of 50 examples for validation.
7. Enter the following link: <http://www.tfca.gob.mx/es/TFCA/Boletin_Laboral>   
   
8. You have to click on all the dates that are marked in blue, the start date is 1 marzo 2021. In this example we will use “1 Octubre 2021”



1. A pdf file will open.



Once this is done, the portal will deliver lists in which we will collect the following data in general:

****

**EXAMPLE (as shown on the Lists):**

1. Fecha: 1 DE OCTUBRE DE 2021
2. Juzgado: PRIMERA SALA
3. Expediente: 5966/12
4. “Actor”: ARCE CORZO ABENAMAR
5. ”Demandado”: SAGARPA
6. Acuerdos: MANIFESTACIONES DE LA PARTE ACTORA

**DICTIONARY**

This is the example of how to fill the dictionary to do the insertion in the ‘raw’ database. We need the data only in upper() and with out any special character ó, ö, etc.:

"actor" : ARCE CORZO ABENAMAR

"demandado" : SAGARPA

"entidad" : CIUDAD DE MEXICO (DEFAULT)

"expediente" :5966/12

"fecha" : 1 DE OCTUBRE DE 2021 (Note: we must indicate the most recent “fecha” of the “Acuerdos” extracted)

"fuero" : FEDERAL (DEFAULT)

"juzgado" : PRIMERA SALA

"tipo" : LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

"acuerdos" : MANIFESTACIONES DE LA PARTE ACTORA

“monto”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“fecha\_presentacion”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“actos\_reclamados”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“actos\_reclamados\_especificos”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“Naturaleza\_procedimiento”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“Prestación\_demandada”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“Organo\_jurisdiccional\_origen”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“expediente\_origen”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“materia”: LABORAL (DEFAULT)

“submateria”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“fecha\_sentencia”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

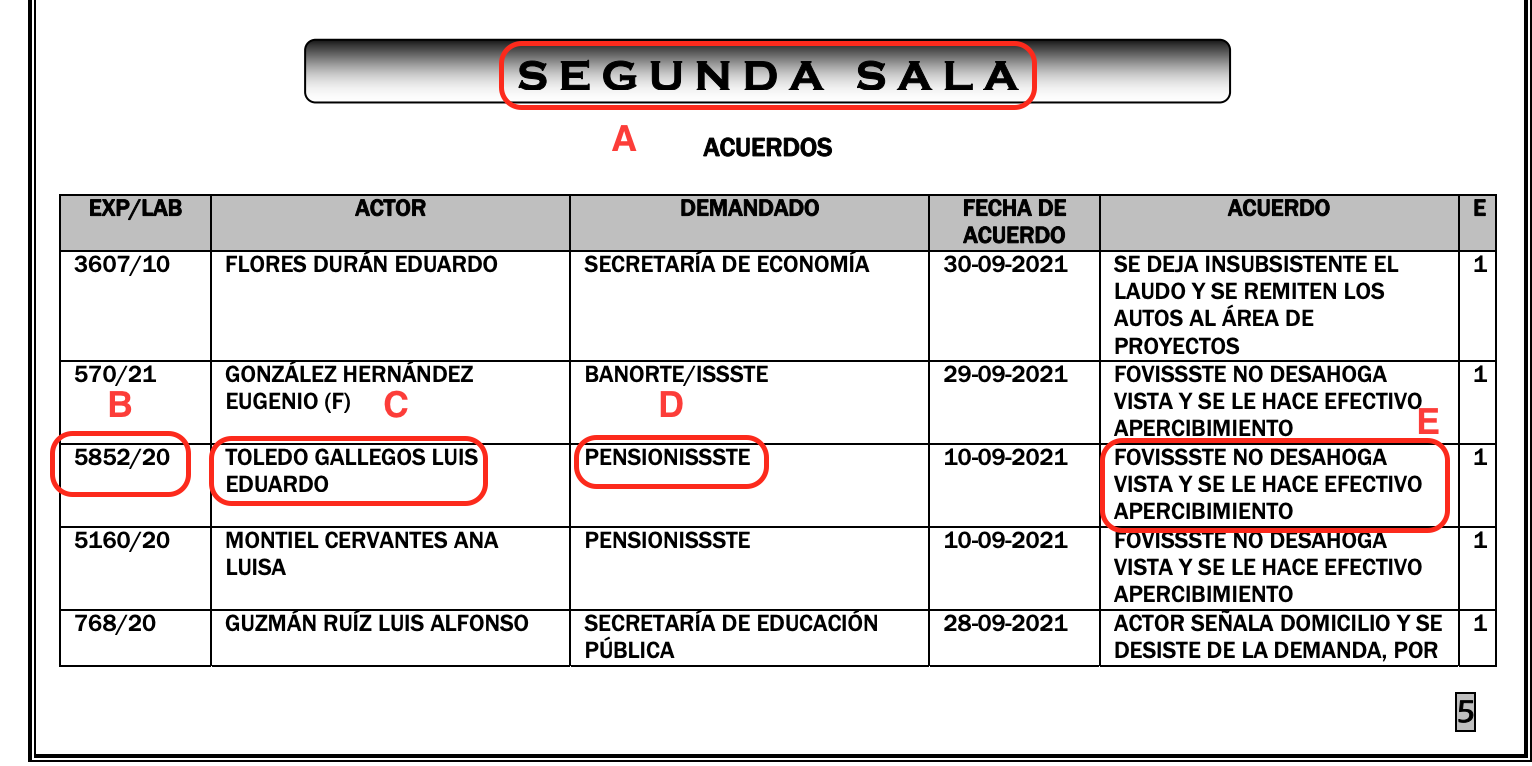
“sentido\_sentencia”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“resoluciones”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“origen: TRIBUNAL FEDERAL DE CONCILIACION Y ARBITRAJE

“Fecha insercion: current time, you can use datetime.now()

Once this is done, the portal will deliver lists in which we will collect the following data in general:

**EXAMPLE (as shown on the Lists):**

1. Juzgado: SEGUNDA SALA
2. Expediente: 5852/20
3. “Actor”: TOLEDO GALLEGOS LUIS EDUARDO
4. ”Demandado”: PENSION/ISSSTE
5. Acuerdos: FOVISSSTE NO DESAOGA VISTA Y SE LE HACE EFECTIVO EL APERCIBIMIENTO.

**DICTIONARY**

"actor" : TOLEDO GALLEGOS LUIS EDUARDO

"demandado" :PENSION/ISSSTE

"entidad" : CIUDAD DE MEXICO

"expediente" : 5852/20

"fecha" : 1 DE OCTUBRE DE 2021 (DATE IT ALWAYS GOING TO BE THE SAME AS THE DATE OF “BOLETIN” (Note: we must indicate the most recent “fecha” of the “Acuerdos” extracted))

"fuero" : FEDERAL (DEFAULT)

"juzgado" : SEGUNDA SALA

"tipo" : LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

"acuerdos" : FOVISSSTE NO DESAOGA VISTA Y SE LE HACE EFECTIVO EL APERCIBIMIENTO.

“monto”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“fecha\_presentacion”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“actos\_reclamados”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“actos\_reclamados\_especificos”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“Naturaleza\_procedimiento”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“Prestación\_demandada”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“Organo\_jurisdiccional\_origen”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“expediente\_origen”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“materia”: LABORAL (DEFAULT)

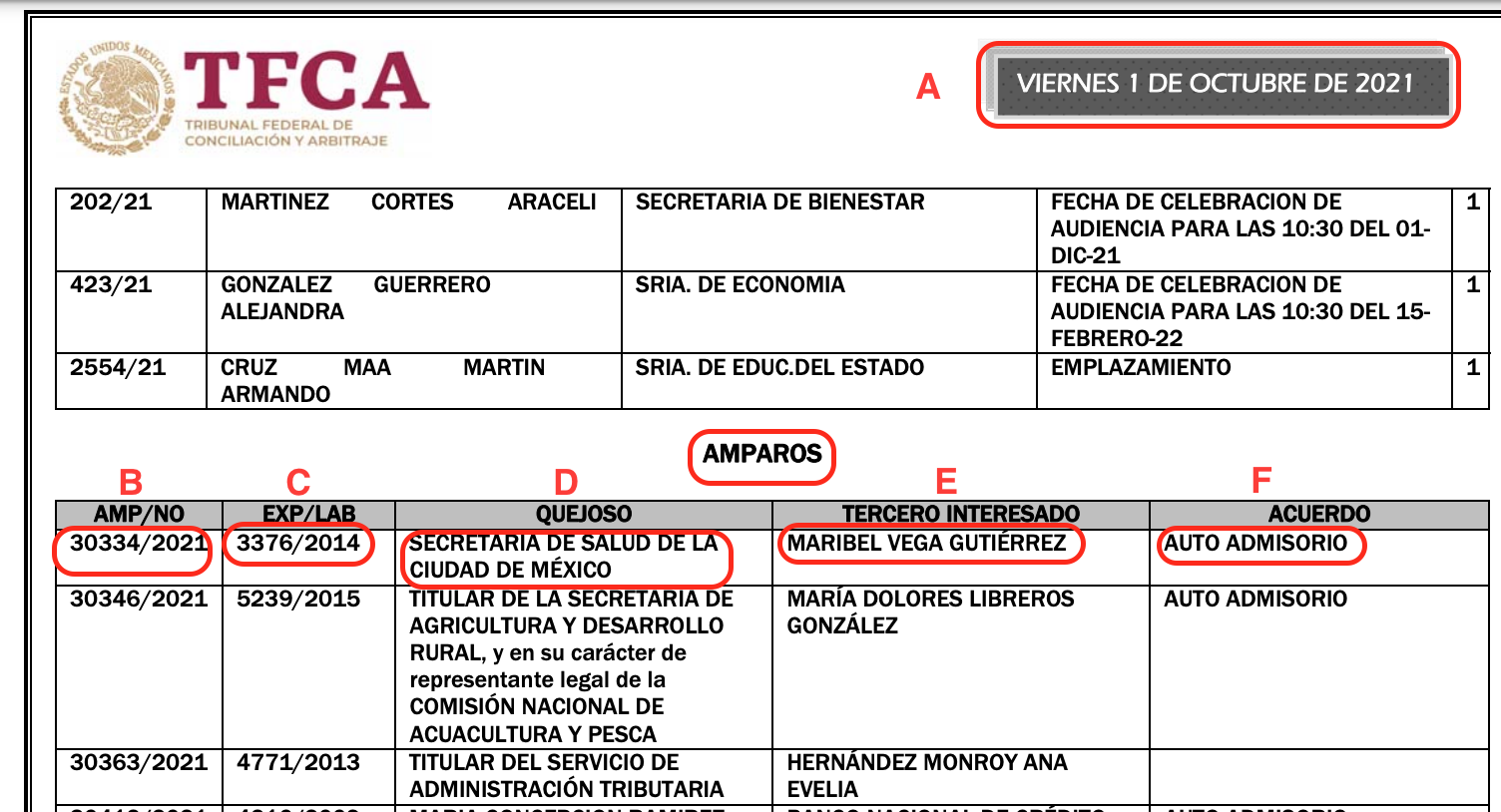
“submateria”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“fecha\_sentencia”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“sentido\_sentencia”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“resoluciones”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“origen: TRIBUNAL FEDERAL DE CONCILIACION Y ARBITRAJE  
“Fecha insercion: datetime.now()



Once this is done, the portal will deliver lists in which we will collect the following data in general:

**EXAMPLE (as shown on the Lists):**

1. Fecha: 1 DE OCTUBRE DE 2021
2. Amparo: 30334/2021
3. Expediente: 3376/2014
4. “Actor”: SECRETARIA DE SALUD DE LA CIUDAD DE MEXICO
5. ”Demandado”: MARIBEL VEGA GUTIERREZ
6. Acuerdos: AUTO ADMISORIO

**DICTIONARY**

{

"actor" : SECRETARIA DE SALUD DE LA CIUDAD DE MEXICO

"demandado" : MARIBEL VEGA GUTIERREZ

"entidad" : CIUDAD DE MEXICO (DEFAULT)

"expediente" : 30334/2021

"fecha" : 1 DE OCTUBRE DE 2021 (DATE IT ALWAYS GOING TO BE THE SAME AS THE DATE OF “BOLETIN” (Note: we must indicate the most recent “fecha” of the “Acuerdos” extracted))

"fuero" :FEDERAL (DEFAULT)

"juzgado" : LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

"tipo" : AMPARO

"acuerdos" : AUTO ADMISORIO

“monto”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“fecha\_presentacion”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“actos\_reclamados”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“actos\_reclamados\_especificos”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“Naturaleza\_procedimiento”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“Prestación\_demandada”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“Organo\_jurisdiccional\_origen”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“expediente\_origen”: 3376/2014

“materia”: LABORAL (DEFAULT)

“submateria”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“fecha\_sentencia”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“sentido\_sentencia”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“resoluciones”: LIKE WE DONT HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

“origen: TRIBUNAL FEDERAL DE CONCILIACION Y ARBITRAJE  
“Fecha insercion: datetime.now()