**TECH SPEC JUNTA LOCAL DE CONCILIACION Y ARBITRAJE VERACRUZ”**

**Code Name: Laboral\_Veracruz**

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

1. **DICTIONARY:** 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"*: ,

*"Prestación\_demandada"*: ,

*"Organo\_jurisdiccional\_origen"*: ,

*"expediente\_origen"*: ,

*"materia"*: ,

*"submateria"*: ,

*"fecha\_sentencia"*: ,

*"sentido\_sentencia"*: ,

*"resoluciones"*: ,

*"origen"*: ,

*"fecha\_insercion"*: ,

*"fecha\_tecnica"*: ,

}

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

- actor

- demandado

- juzgado

- expediente

- materia

- fecha

1. **TECHNICAL SPECIFICATIONS:**

* **Programming language**: Python Request or Python Selenium
* **Dict In Mongodb:** The code must be prepared to insert the data obtained and assigned in the Raw dictionary directly as a dict in MongoDB

1. **PAGE SATURATION:** In case of page saturation, the code must detect it and put it in time sleep for two hours and resume after this time.
2. **INDEX:** 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.
3. **DATA TO TODAY’S DATE:** When the scraper obtains all the existing data to the current date (today’s date), it will have to wait 48 hours to extract new data again.
4. **START-DATE AND END-DATE:** 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 the end-date is today and scrap until today and then day by day to keep extracting new available data.
5. **SAMPLES:** Once the scraper is working, you must send samples in json format with a minimum of 50 examples for validation.
6. **LINK:** Enter the following link: <http://www.veracruz.gob.mx/trabajo/junta-local-de-conciliacion-y-arbitraje/>
7. **MAIN PAGE:** Select the option “Lista de Acuerdos de las Juntas Especiales de la Local de Conciliacion y Arbitraje Del Estado”. It will show us the “Lista de acuerdos”, from where the information of the entire list that is displayed and of those that are added to the page must be extracted



**1**

**2**



**A**

1. **BULLETIN START-DATE:** When selecting a “Junta Local”. It will show us a calendar to select a day, which will be considered the publication date, must be saved in format: yyyy/mm/dd. 6 January 2020 should be considered as the start date for data extraction.



**B**

1. **PDF:** Once you have selected the corresponding date, you will find the PDF that contains the scanned page with the information to extract.

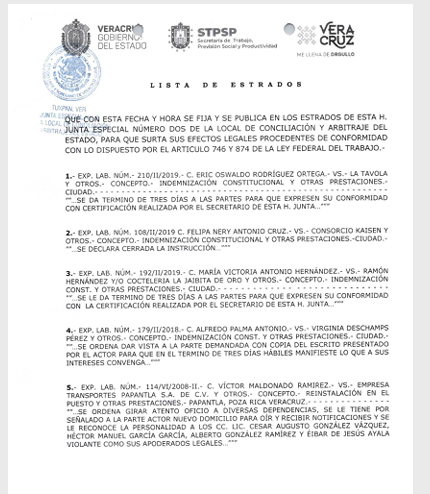
NOTE: To avoid overloading the website, we need to work in two blocks:

1. Information Collection: The scraper must go to the website to collect and/or download the PDF's from the start date to the most updated date, storing them in the VM.

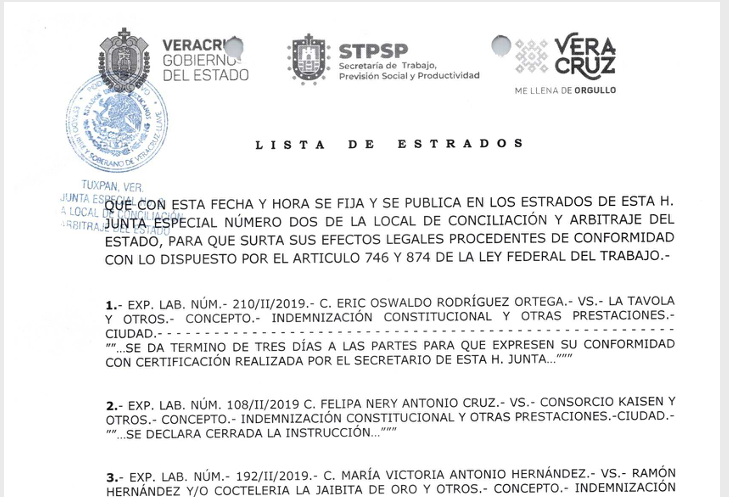
2. The Scraper must go back to storage to read the PDF's and process the information as requested below.

Pursuant to the above, the next time you run it, you only need to check the latest pdf and if there is a new pdf.

Hence the trial and error process of the OCR/data extraction from the pdf is done within the VM itself



1. **EXAMPLE I:** Considering the previous steps, the following example of the relevant information to be collected is shown:



**D**

**G**

**F**

**E**

**C**

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

1. **JUZGADO:** JUNTA ESPECIAL NUMERO DOS DE LA LOCAL DE CONCILIACON Y ARBITRAJE DEL ESTADO. (It will be the one that was selected on point XI)
2. **FECHA:** 6 DE ENERO DE 2020 (It will be the date that was selected on point XII)
3. **EXPEDIENTE:** 210/II/2019
4. **ACTOR:** OSWALDO RODRIGUEZ ORTEGA
5. **DEMANDADO:** LA TAVOLA Y OTROS.
6. **TIPO:** INDEMNIZACION CONSTITUCIONAL Y OTRAS PRESTACIONES.
7. **ACUERDOS:** SE DA TERMINO DE TRES DIAS A LAS PARTES PARA QUE EXPRESEN SU CONFORMIDAD CON CERTIFICACION REALIZADA POR EL SECRETARIO DE ESTA H. JUNTA…

**DICTIONARY**

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

{

*"actor"* : " OSWALDO RODRIGUEZ ORTEGA ",

*"demandado"* : " LA TAVOLA Y OTROS ",

*"entidad"* : "VERACRUZ", (BY DEFAULT)

*"expediente"* : "210/II/2019",

*"fecha"* : "2020/01/06", (It will be the date that was selected on point "XII" (Note: we must indicate the most recent "fecha" of the "Acuerdos" extracted))

*"fuero"* : "COMUN",

*"juzgado"*: "JUNTA ESPECIAL NUMERO DOS DE LA LOCAL DE CONCILIACON Y ARBITRAJE DEL ESTADO.",

*"tipo"* : : “INDEMNIZACION CONSTITUCIONAL Y OTRAS PRESTACIONES.”,

*"acuerdos"* : " SE DA TERMINO DE TRES DIAS A LAS PARTES PARA QUE EXPRESEN SU CONFORMIDAD CON CERTIFICACION REALIZADA POR EL SECRETARIO DE ESTA H. JUNTA…",

*"monto"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_presentacion"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados\_especificos"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"naturaleza\_procedimiento"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"prestación\_demandada"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"organo\_jurisdiccional\_origen"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"expediente\_origen"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"materia"*: "LABORAL", (DEFAULT)

*"submateria"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"sentido\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"resoluciones"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"origen"*: " JUNTAS ESPECIALES DE CONCILIACION Y ARBITRAJE DEL ESTADO DE VERACRUZ", (BY DEFAULT)

*"fecha\_insercion"*: current time\*, you can use datetime.now() insert as ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

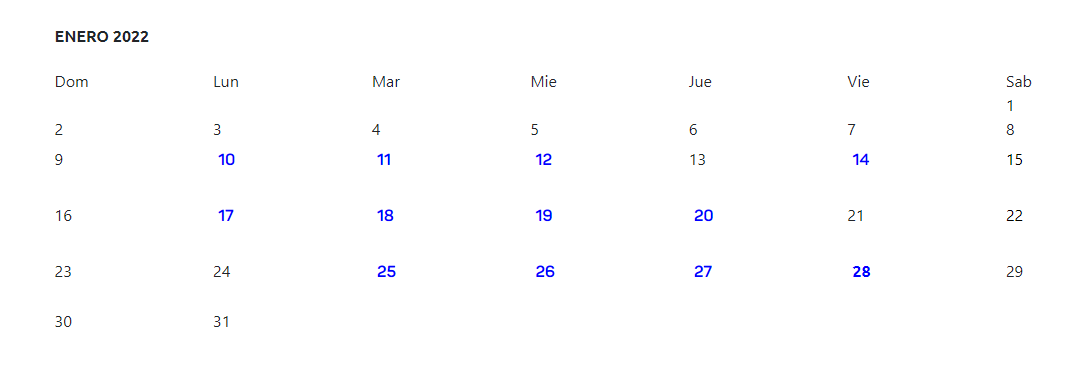
*"fecha\_tecnica"*: value of *"date"* in ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

}

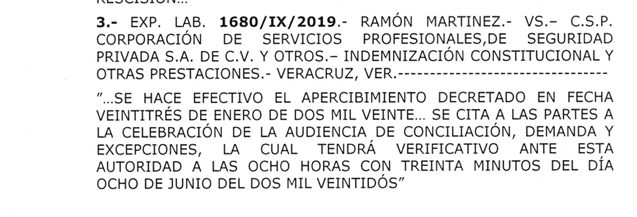
1. **Example II:**



**A**



**B**



**G**

**F**

**E**

**D**

**C**

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

1. **JUZGADO:** JUNTA ESPECIAL NUMERO NUEVE DE LA LOCAL DE CONCILIACON Y ARBITRAJE DEL ESTADO. (It will be the one that was selected on point XI)
2. **FECHA:** 26 DE ENERO DE 2022 (It will be the date that was selected on point XII)
3. **EXPEDIENTE:** 1680/IX/2019
4. **ACTOR:** RAMON MARTINEZ
5. **DEMANDADO:** C.S.P CORPORACION DE SERVICIOS PROFESIONALES, DE SEGURIDAD PRIVADA S.A. DE C.V. Y OTROS.
6. **TIPO:** INDEMNIZACION CONSTITUCIONAL Y OTRAS PRESTACIONES.
7. **ACUERDOS:** SE HACE EFECTIVO EL APERCIVIMIENTO DECRETADO EN FECHA VEINTITRÉS DE ENERO DE DOS MIL VEINTE… SE CITA A LAS PARTES A LA CELEBRACION DE LA AUDIENCIA DE CONCILIACION, DEMANDA Y EXCEPCIONES, LA CUAL TENDRA VERIFICATIVO ANTE ESTA AUTORIDAD A LAS OCHO HORAS CON TREINTA MINUTOS DEL DIA OCHO DE JUNIO DE DOS MIL VEINTIDOS.

**DICTIONARY**

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

{

*"actor"* : " RAMON MARTINEZ ",

*"demandado"* : " C.S.P CORPORACION DE SERVICIOS PROFESIONALES, DE SEGURIDAD PRIVADA S.A. DE C.V. Y OTROS.",

*"entidad"* : "VERACRUZ", (BY DEFAULT)

*"expediente"* : "1680/IX/2019",

*"fecha"* : "2022/01/26", (It will be the date that was selected on point "XII" (Note: we must indicate the most recent "fecha" of the "Acuerdos" extracted))

*"fuero"* : "COMUN",

*"juzgado"*: "JUNTA ESPECIAL NUMERO NUEVE DE LA LOCAL DE CONCILIACON Y ARBITRAJE DEL ESTADO.",

*"tipo"* : : “INDEMNIZACION CONSTITUCIONAL Y OTRAS PRESTACIONES.”,

*"acuerdos"* : " SE HACE EFECTIVO EL APERCIVIMIENTO DECRETADO EN FECHA VEINTITRÉS DE ENERO DE DOS MIL VEINTE… SE CITA A LAS PARTES A LA CELEBRACION DE LA AUDIENCIA DE CONCILIACION, DEMANDA Y EXCEPCIONES, LA CUAL TENDRA VERIFICATIVO ANTE ESTA AUTORIDAD A LAS OCHO HORAS CON TREINTA MINUTOS DEL DIA OCHO DE JUNIO DE DOS MIL VEINTIDOS. …",

*"monto"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_presentacion"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados\_especificos"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"naturaleza\_procedimiento"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"prestación\_demandada"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"organo\_jurisdiccional\_origen"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"expediente\_origen"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"materia"*: "LABORAL", (DEFAULT)

*"submateria"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"sentido\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"resoluciones"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"origen"*: " JUNTAS ESPECIALES DE CONCILIACION Y ARBITRAJE DEL ESTADO DE VERACRUZ", (BY DEFAULT)

*"fecha\_insercion"*: current time\*, you can use datetime.now() insert as ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

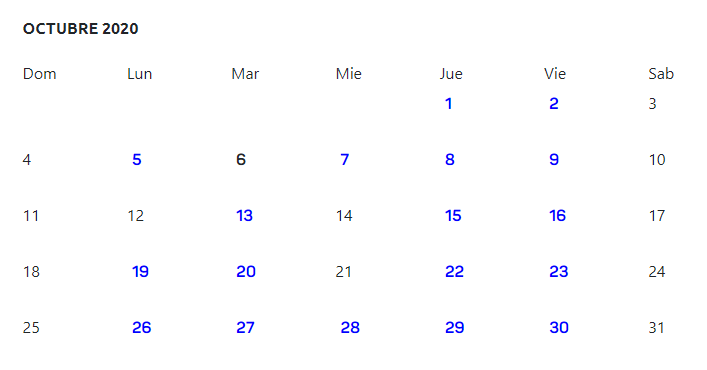
*"fecha\_tecnica"*: value of *"date"* in ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

}

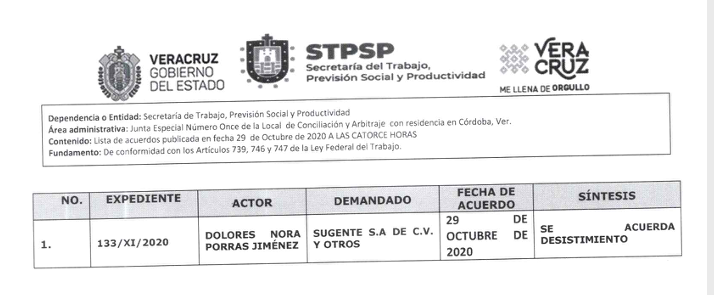
1. Example III: You can find a different kind of document:



**A**



**B**



**F**

**E**

**D**

**C**

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

1. **JUZGADO:** JUNTA ESPECIAL NUMERO ONCE DE LA LOCAL DE CONCILIACON Y ARBITRAJE DEL ESTADO. (It will be the one that was selected on point XI)
2. **FECHA:** 29 DE OCTUBRE DE 2020 (It will be the date that was selected on point XII)
3. **EXPEDIENTE:** 133/XI/2020
4. **ACTOR:** DOLORES PORRA NORA JIMENEZ
5. **DEMANDADO:** SUGENTE S.A DE C.V Y OTROS
6. **ACUERDOS:** SE ACUERDA DESISTIMIENTO

**DICTIONARY**

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

{

*"actor"* : " DOLORES PORRA NORA JIMENEZ ",

*"demandado"* : "SUGENTE S.A DE C.V Y OTROS ",

*"entidad"* : "VERACRUZ", (BY DEFAULT)

*"expediente"* : "133/XI/2020",

*"fecha"* : "2020/10/29", (It will be the date that was selected on point "XII" (Note: we must indicate the most recent "fecha" of the "Acuerdos" extracted))

*"fuero"* : "COMUN",

*"juzgado"*: "JUNTA ESPECIAL NUMERO ONCE DE LA LOCAL DE CONCILIACON Y ARBITRAJE DEL ESTADO.",

*"tipo"* : :, LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"acuerdos"* : " SE ACUERDA DESISTIMIENTO ",

*"monto"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_presentacion"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados\_especificos"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"naturaleza\_procedimiento"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"prestación\_demandada"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"organo\_jurisdiccional\_origen"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"expediente\_origen"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"materia"*: "LABORAL", (DEFAULT)

*"submateria"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"sentido\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"resoluciones"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

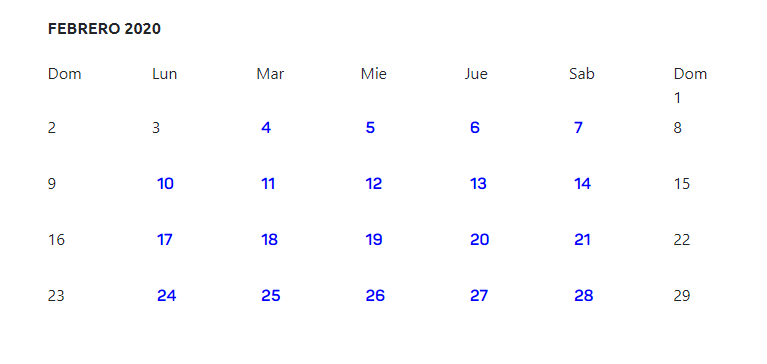
*"origen"*: " JUNTAS ESPECIALES DE CONCILIACION Y ARBITRAJE DEL ESTADO DE VERACRUZ", (BY DEFAULT)

*"fecha\_insercion"*: current time\*, you can use datetime.now() insert as ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

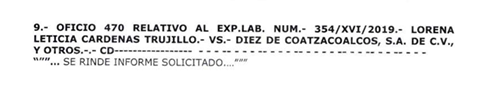
*"fecha\_tecnica"*: value of *"date"* in ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

}

1. Example IV: You can find another kind of data, like this:



**A**



**F**

**E**

**D**

**C**

**B**

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

1. **JUZGADO:** JUNTA ESPECIAL NUMERO DIECISEIS DE LA LOCAL DE CONCILIACON Y ARBITRAJE DEL ESTADO. (It will be the one that was selected on point XI)
2. **FECHA:** 19 DE FEBRERO DE 2020 (It will be the date that was selected on point XII)
3. **EXPEDIENTE:** 354/XVI/2019
4. **ACTOR:** LORENA LETICIA CARDENAS TRUJILLO
5. **DEMANDADO:** DIEZ DE COATZACOALCOS S.A DE C.V
6. **ACUERDOS:** SE RINDE INFORME SOLICITADO

**DICTIONARY**

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

{

*"actor"* : “LORENA LETICIA CARDENAS TRUJILLO ",

*"demandado"* : " DIEZ DE COATZACOALCOS S.A DE C.V ",

*"entidad"* : "VERACRUZ", (BY DEFAULT)

*"expediente"* : "354/XVI/2019",

*"fecha"* : "2020/02/19", (It will be the date that was selected on point "XII" (Note: we must indicate the most recent "fecha" of the "Acuerdos" extracted))

*"fuero"* : "COMUN",

*"juzgado"*: "JUNTA ESPECIAL NUMERO DIECISEIS DE LA LOCAL DE CONCILIACON Y ARBITRAJE DEL ESTADO.",

*"tipo"* : :, LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"acuerdos"* : " SE RINDE INFORME SOLICITADO",

*"monto"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_presentacion"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados\_especificos"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"naturaleza\_procedimiento"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"prestación\_demandada"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"organo\_jurisdiccional\_origen"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"expediente\_origen"*: "", LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT

*"materia"*: "LABORAL", (DEFAULT)

*"submateria"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"sentido\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"resoluciones"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"origen"*: " JUNTAS ESPECIALES DE CONCILIACION Y ARBITRAJE DEL ESTADO DE VERACRUZ", (BY DEFAULT)

*"fecha\_insercion"*: current time\*, you can use datetime.now() insert as ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

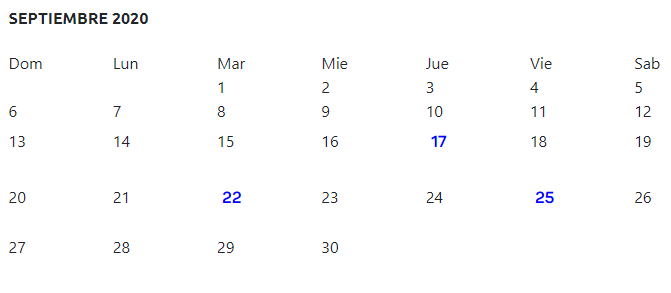
*"fecha\_tecnica"*: value of *"date"* in ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

}

1. Example IV: You can find another kind of data, like this:



**A**



**B**



**E**

**D**

**C**

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

1. **JUZGADO:** DEPARTAMENTO DE CONTRATOS COLECTIVOS DE TRABAJO (It will be the one that was selected on point XI)
2. **FECHA:** 17 DE SEPTIEMBRE DE 2020 (It will be the date that was selected on point XII)
3. **EXPEDIENTE:** CCI-212/2019
4. **ACTOR:** COMPAÑIA MEXICANA DE TRASLADO DE VALORES S.A. DE C.V.
5. **DEMANDADO:** SINDICATO NACIONAL DE TRABAJADORES TECNICOS Y EMPLEADOS DE LA INDUSTRIA DEL TRANSPORTE Y DE LA CONSTRUCCION SIMILARES Y CONEXOS DE LA REPUBLICA MEXICANA

**DICTIONARY**

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

{

*"actor"* : “COMPAÑIA MEXICANA DE TRASLADO DE VALORES S.A. DE C.V.",

*"demandado"* : " SINDICATO NACIONAL DE TRABAJADORES TECNICOS Y EMPLEADOS DE LA INDUSTRIA DEL TRANSPORTE Y DE LA CONSTRUCCION SIMILARES Y CONEXOS DE LA REPUBLICA MEXICANA ",

*"entidad"* : "VERACRUZ", (BY DEFAULT)

*"expediente"* : " CCI-212/2019 ",

*"fecha"* : "2020/09/17", (It will be the date that was selected on point "XII" (Note: we must indicate the most recent "fecha" of the "Acuerdos" extracted))

*"fuero"* : "COMUN",

*"juzgado"*: " DEPARTAMENTO DE CONTRATOS COLECTIVOS DE TRABAJO ",

*"tipo"* : :, LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"acuerdos"* : "", LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"monto"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_presentacion"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"actos\_reclamados\_especificos"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"naturaleza\_procedimiento"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"prestación\_demandada"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"organo\_jurisdiccional\_origen"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"expediente\_origen"*: "", LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"materia"*: "LABORAL", (DEFAULT)

*"submateria"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"fecha\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"sentido\_sentencia"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"resoluciones"*: "",LIKE WE DON’T HAVE DATA TO FILL THIS FIELD, WE NEED TO PUT IT LIKE EMPTY

*"origen"*: " JUNTAS ESPECIALES DE CONCILIACION Y ARBITRAJE DEL ESTADO DE VERACRUZ", (BY DEFAULT)

*"fecha\_insercion"*: current time\*, you can use datetime.now() insert as ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

*"fecha\_tecnica"*: value of *"date"* in ISOdate for mongo db and datetime for python (not text / str) \*\*\*IMPORTANT\*\*\*

}

[7:56 PM, 5/19/2022] Irshad Khan Freelance: Needed right away, with long-term possibility as we have constant and non-stop projects for the next 12 months. We are a team working with the best freelance developers creating and maintaining long-term profitable relationships (not necessary if not interested). We are reliable, trustworthy, and friendly recruiters; we've been hiring freelancers since the conception of the start-up.

What we're looking for:

a. Top expert Python experience mandatory (Python Request and/or Selenium), other languages welcome but not necessary

b. Strong web scraping and pdf/images data extraction (OCR)

c. Expert on bypassing website security protection: fakeuser, proxies, captchas, etc.

d. Maximum confidentiality, no disclosing of any work done with us

[7:57 PM, 5/19/2022] Irshad Khan Freelance: Hello!

We’d like to introduce ourselves so you get an idea of who we are and what we’re looking for.

First of all, we are a start-up based in Mexico that performs background checks. We currently operate with 20+ freelancers worldwide on a non-stop project basis, meaning work is active and abundant if that is what you are looking for.

The projects mainly consist of scraping bots with Python Request/Selenium (depending on the website). The process is as follows:

1. Scrap the website

2. Avoid security

3. Download PDF canvas/images

4. Extract necessary data from the DPF/Canvas/images with algorithm/OCR

5. Insert data extracted as a dictionary on Mongo DB

We have over 100 projects ready to assign. What we need from you is proof that you have the necessary knowledge and persistence to partake in these projects, while being able to execute them successfully in a timeframe of 10 days or less. We need constant communication and updates at least every 3 to 4 days to understand how you are progressing, and to know whether you will be able to deliver the project on time.

We have clear lines drawn on the sand:

1. Projects' budget vary based on complexity from USD 500 all the way up to USD 1,000 in certain cases

2. We pay only after samples have been delivered

3. After 10 days if samples have not been delivered we will understand you are not capable of working on the project and proceed to eliminate you from the team — we are looking for excellence, resilience, persistence, and strong intensity of work

4. Lastly, if you deliver successfully, we guarantee an extremely profitable long-term relation with constant projects for at least 12 months