**TECH SPEC SECRETARIA DEL TRABAJO ESTADO DE COAHUILA”**

**Code Name:** Laboral\_Coahuila

**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. **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.
2. **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.
3. **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.
4. **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.
5. **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.
6. **SAMPLES:** Once the scraper is working, you must send samples in json format with a minimum of 50 examples for validation.
7. **LINK:** Enter the following link: <http://setracoahuila.gob.mx/>
8. Scroll down and then select the option “Justicia Laboral”, then the option “Consulta de Resoluciones”



1. The last step will lead us to the section that has the following links, this tech spec will appy to them:



**Each of the previous sections will correspond to the “juzgado” field, for example:**

<http://187.160.245.251:8088/acuerdos/Listasaltillo.jsp> (JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE SALTILLO)

*"juzgado"*: "JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE SALTILLO",



(EXAMPLE)

<http://187.160.245.251:8088/acuerdos/Listasaltilloe.jsp> (JUNTA ESPECIAL DE CONCILIACION Y ARBITRAJE DE SALTILLO)

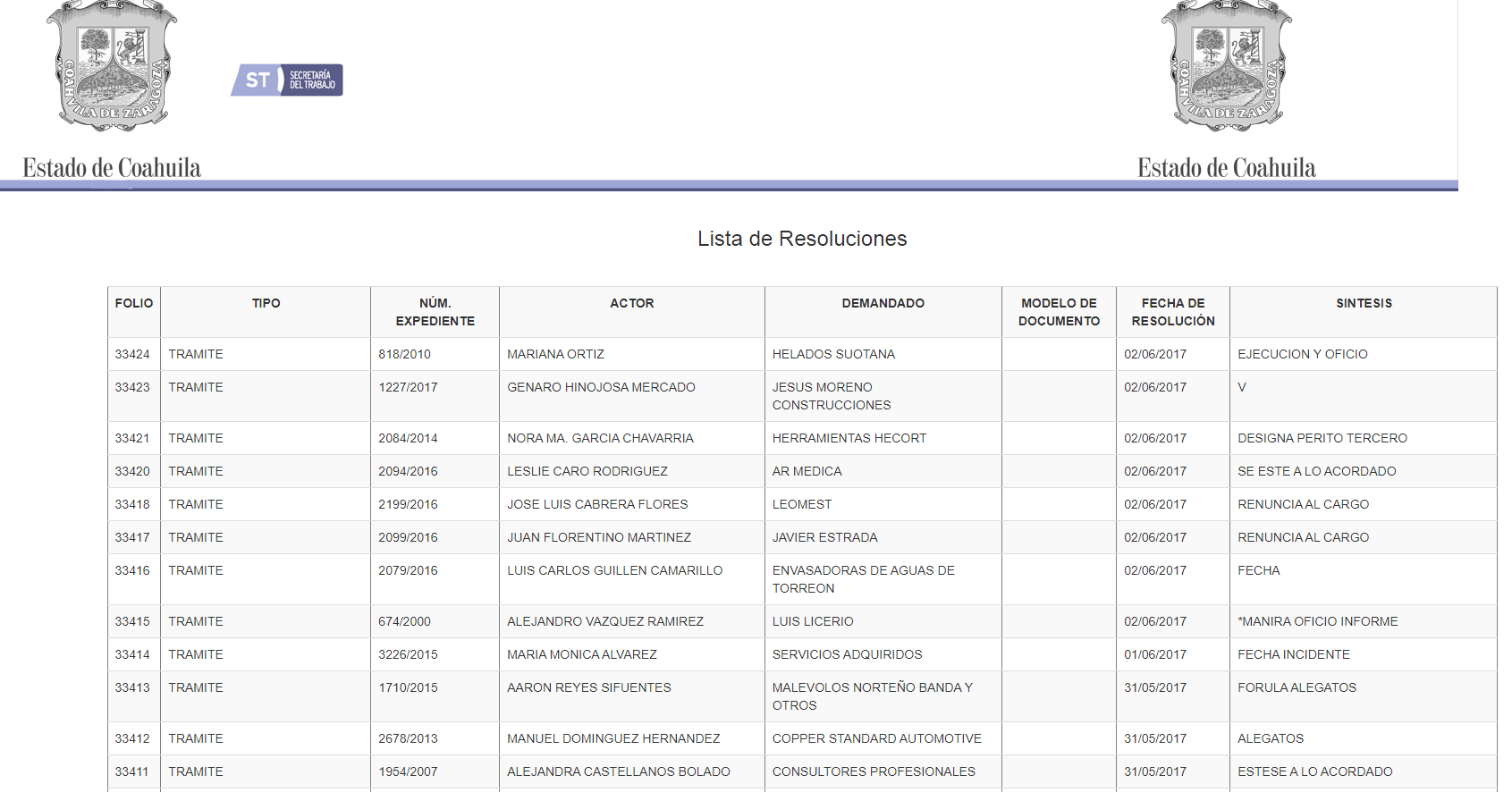
*"juzgado"*: " JUNTA ESPECIAL DE CONCILIACION Y ARBITRAJE DE SALTILLO",



(EXAMPLE)

<http://187.160.245.251:8088/acuerdos/Listatorreon.jsp> (JUNTA LOCAL DE DE CONCILIACION Y ARBITRAJE DE CONCILIACION Y ARBITRAJE DE TORREON)

*"juzgado"*: " JUNTA LOCAL DE DE CONCILIACION Y ARBITRAJE DE CONCILIACION Y ARBITRAJE DE TORREON",



(EXAMPLE)  
  
<http://187.160.245.251:8088/acuerdos/Listamonclova.jsp> (JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE MONCLOVA)

*"juzgado"*: " JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE MONCLOVA ",

(EXAMPLE)

<http://187.160.245.251:8088/acuerdos/Listapiedras.jsp> (JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE PIEDRAS NEGRAS)

*"juzgado"*: " JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE PIEDRAS NEGRAS",



(EXAMPLE)

<http://187.160.245.251:8088/acuerdos/Listasabinas.jsp> (JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE SABINAS)

*"juzgado"*: " JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE SABINAS",



(EXAMPLE)

<http://187.160.245.251:8088/acuerdos/Listaacuna.jsp> (JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE ACUÑA)

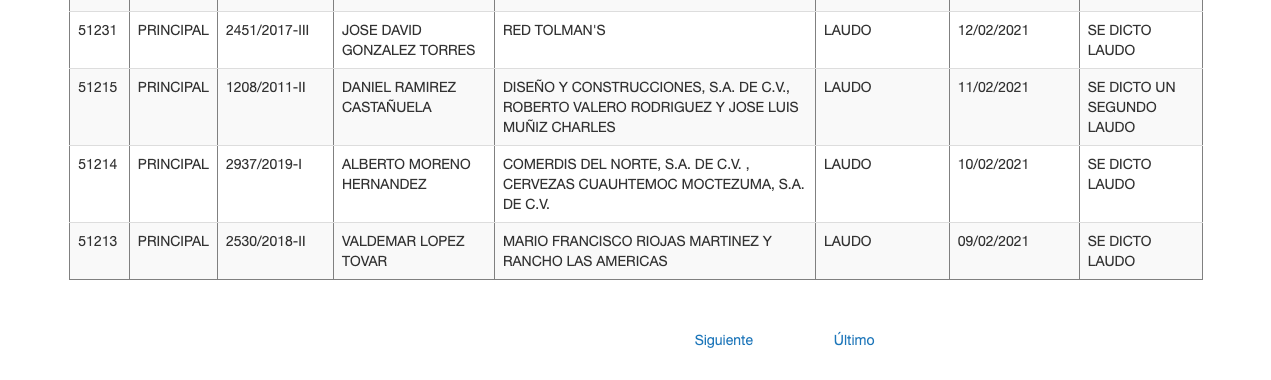
*"juzgado"*: "JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE ACUÑA",



(EXAMPLE)

1. Extract the information of each link from the first page to the last.





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

**Example 1 Junta Especial Saltillo**



**D**

V

**F**

**E**

**C**

**B**

**A**

V

1. EXPEDIENTE : 1952/2016-IV (EXAMPLE)
2. ACTOR: JESUS GUADALUPE PINTOR LUCIO
3. DEMANDADO: YONG SAN AUTOMOTIVE MEXICO S DE RL DE CV
4. TIPO : PRINCIPAL
5. ACUERDO: SE DICTO LAUDO
6. FECHA: 07/03/2021

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 without any special character ó, ö, etc.:

{

*"actor"* : " JESUS GUADALUPE PINTOR LUCIO",

*"demandado"* : " YONG SAN AUTOMOTIVE MEXICO S DE RL DE CV ",

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

*"expediente"* : "1952/2016-IV ",

*"fecha"* : "2021/03/07", (must be saved in format: yyyy/mm/dd)

*"fuero"* : "COMUN",

*"juzgado"*: " JUNTA ESPECIAL SALTILLO ", (This data will be by default according to point XII, each link is a “JUNTA”)

*"tipo"* : "PRINCIPAL",

*"acuerdos"* : "SE DICTO LAUDO",

*"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",

*"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

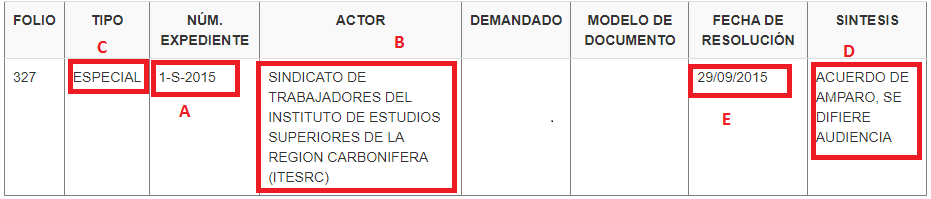
*"origen"*: " SECRETARIA DEL TRABAJO DEL ESTADO DE COAHUILA ",

*"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\*\*\*

}

**Example 2 Junta Local de Conciliación y Arbitraje de Monclova**



1. EXPEDIENTE : 1-S-2015 (EXAMPLE)
2. ACTOR: SINDICATO DE TRABAJADORES DEL INSTITUTO DE ESTUDIOS SUPERIORES DE LA REGION CARBONIFERA (ITESRC)
3. TIPO : ESPECIAL
4. ACUERDO: ACUERDO DE AMPARO, SE DIFIERE AUDIENCIA
5. FECHA: 29/09/2015

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 without any special character ó, ö, etc.:

{

*"actor"* : "SINDICATO DE TRABAJADORES DEL INSTITUTO DE ESTUDIOS SUPERIORES DE LA REGION CARBONIFERA (ITESRC)",

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

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

*"expediente"* : "1-S-2015 ",

*"fecha"* : "2015/09/29", (must be saved in format: yyyy/mm/dd)

*"fuero"* : "COMUN",

*"juzgado"*: " JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE MONCLOVA", (This data will be by default according to point XII, each link is a “JUNTA”)

*"tipo"* : " ESPECIAL",

*"acuerdos"* : " ACUERDO DE AMPARO, SE DIFIERE AUDIENCIA ",

*"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",

*"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

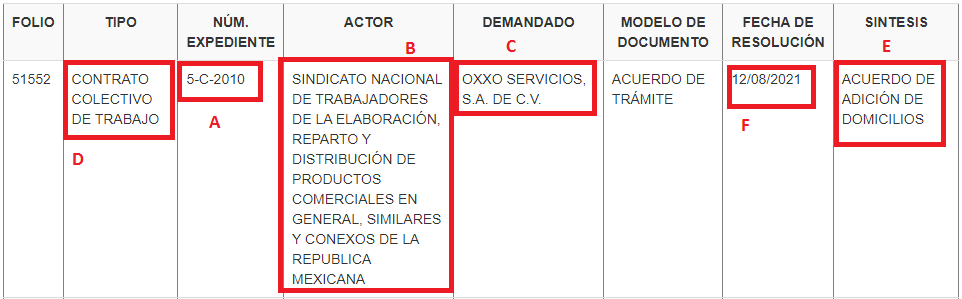
*"origen"*: " SECRETARIA DEL TRABAJO DEL ESTADO DE COAHUILA ",

*"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\*\*\*

}

**Example 3 Junta Local de Conciliación y Arbitraje de Sabinas**



1. EXPEDIENTE : 5-C-2010 (EXAMPLE)
2. ACTOR: SINDICATO NACIONAL DE TRABAJADORES DE LA ELABORACIÓN, REPARTO Y DISTRIBUCIÓN DE PRODUCTOS COMERCIALES EN GENERAL, SIMILARES Y CONEXOS DE LA REPUBLICA MEXICANA
3. DEMANDADO: OXXO SERVICIOS, S.A. DE C.V.
4. TIPO : CONTRATO COLECTIVO DE TRABAJO
5. ACUERDO: ACUERDO DE ADICIÓN DE DOMICILIOS
6. FECHA: 12/08/2021

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 without any special character ó, ö, etc.:

{

*"actor"* : " SINDICATO NACIONAL DE TRABAJADORES DE LA ELABORACIÓN, REPARTO Y DISTRIBUCIÓN DE PRODUCTOS COMERCIALES EN GENERAL, SIMILARES Y CONEXOS DE LA REPUBLICA MEXICANA",

*"demandado"* : " OXXO SERVICIOS, S.A. DE C.V.",

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

*"expediente"* : "5-C-2010 ",

*"fecha"* : "2021/08/12", (must be saved in format: yyyy/mm/dd)

*"fuero"* : "COMUN",

*"juzgado"*: " JUNTA LOCAL DE CONCILIACION Y ARBITRAJE DE SABINAS", (This data will be by default according to point XII, each link is a “JUNTA”)

*"tipo"* : "PRINCIPAL",

*"acuerdos"* : " ACUERDO DE ADICIÓN DE DOMICILIOS",

*"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",

*"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"*: " SECRETARIA DEL TRABAJO DEL ESTADO DE COAHUILA ",

*"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\*\*\*

}