**ESTRUCTURA DELS REGISTRES A MONGODB**

**Nom de la *bbdd***: “congres”

***Collections* d’aquesta bbdd:**

|  |  |
| --- | --- |
| **Collection** | **Descripció** |
| u'legislatures' |  |
| u'all\_document' |  |
| **u'document'** | Collection amb els documents separats per preguntes i intervencions per diputats |
| u'congres\_members' | Diputats del parlament |
| u'congres\_groups' | Grups polítics al parlament |

**Estructura del registres (documents) a la collection “document”:**

|  |  |
| --- | --- |
| **Claus** | **Descripció** |
| u'session\_dictionary' | llista de diccionaris amb les preguntes i intervencions de la sessió de control |
| u'content\_doc' | document sencer en html |
| u'description' | descripció del document |
| u'url' | url d’extracció del document |
| u'updated\_content\_doc' | data d’última actualització del content\_doc |
| u'date' | data del document (Diario de sesiones del congreso de los diputados) |
| u'updated\_session\_dictionary' | data d’última actualització del session\_dictionary |
| u'\_id' | clau única (creada per MongoDB) |

**Estructura dels diccionaris de la llista u’session\_dictionary’**

Cada element de la llista és un diccionari amb la següent estructura:

|  |  |
| --- | --- |
| **Claus** | **Descripció** |
| u'question’ | Pregunta del diputat |
| u’speaker’ | Diputat que fa la pregunta |
| u'intervention\_dictionary’ | Llista de diccionaris amb les intervencions |
| u'group’ | Grup parlamentari del diputat que fa la pregunta |
| u'num\_exp’ | Número d’expedient de la pregunta |
| u’keywords’ | Llista de tuples (keyword, score) |
| u’numWords’ | Número de paraules de la *question* |
| u’numLines’ | Número de linies de la *question* |

**Estructura dels diccionaris de la llista u'intervention\_dictionary’**

Cada element de la llista és un diccionari amb la següent estructura:

|  |  |
| --- | --- |
| **Claus** | **Descripció** |
| u’who’ | Diccionari de qui fa la intervenció. Les claus són:   * u’name’ * u’surname’ * u’group’ |
| u’text’ | Texte raw de la pregunta / intervenció |
| u’keywords’ | Llista de tuples (keyword, score) |
| u’numWords’ | Número de paraules de *text* |
| u’numLines’ | Número de linies de *text* |