

Motor de Búsqueda



Antonio Abad Hernández Gálvez

Motores de Búsqueda
Universidad de Huelva

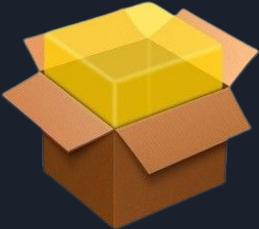




ÍNDICE

- **Organización del Programa**
- **Mejoras del Motor de Búsqueda**
 - Mejora 1: Tipo del campo “texto”
 - Mejora 2: Lista de palabras vacías (StopWords)
 - Comparativa de Resultados
 - Todos los documentos
 - 30 documentos
- **Ejecución**
 - Búsqueda por texto
 - Búsqueda por texto (sin sentido)
 - Búsqueda por id

Organización del Programa



Paquete PRÁCTICA

- Clase **Principal** → Ejecución de las diferentes versiones
- Clase **Separadora** → Corpus Medline y muestreo por consola
- Clase **Fichero** → Manipulación de Ficheros
- Clase **Solr** → Comunicación con Apache Solr
- Clase **Rag** → Genera el JSON de respuesta del RAG



Organización del Programa



Paquete Gráfico (GUI)

- **VistaBuscador** → Interfaz Gráfica Swing



Organización del Programa

```
# pip install -qU langchain[anthropic]
from langchain.agents import create_agent

def get_weather(city: str) -> str:
    """Get weather for a given city."""
    return f"It's always sunny in {city}!"

agent = create_agent(
    model="claude-sonnet-4-5-20250929",
    tools=[get_weather],
    system_prompt="You are a helpful assistant",
)
```

Paquete RAG

- **ConsultaRag** → Consulta y Tratamiento de la respuesta del RAG



Mejoras del Motor de Búsqueda

Mejora 1

- Sobre el campo ‘texto’ de la colección Solr

Pantalla de Esquema

The screenshot shows the Solr schema configuration interface. On the left, there's a sidebar with navigation links like Dashboard, Logging, Security, Core Admin, Java Properties, Thread Dump, and a dropdown menu for 'coleccional inicial'. The main area displays the schema definition for the 'texto' field:

- Field:** texto
- Copied to:** texto_str
- Type:** text_general
- delete field** button (with a red arrow pointing to it from the text below)

Below the field definition, there's a section for Unique Key Field with 'id' listed. Under Global Similarity, it says 'SchemaSimilarity, Default: BM25(k1=1.2,b=0.75)'. At the bottom, there are 'Segments info' and 'Schema' buttons.

Estado inicial del campo

This screenshot shows the 'Manipulate Field Type' interface for the 'texto' field. It includes the following details:

- Field-Type:** org.apache.solr.schema.TextField
- PL.Gap:** 100
- Docs:** 1,033
- Flags:** Indexed, Tokenized, Stored, Multivalued (all checked)
- Properties:** checked
- Schema:** checked
- Index:** checked
- Index Analyzer:** org.apache.solr.analysis.TOKENIZERCHAIN
- Query Analyzer:** org.apache.solr.analysis.TOKENIZERCHAIN
- Load Term Info:** 10 / 13,614 Top-Terms:
 - 1027 of
 - 1021 the
 - 991 and
 - 988 in
 - 867 a
 - 857 to
 - 760 with
 - 637 is
 - 600 by
 - 563 wasHistogram showing term frequency counts.

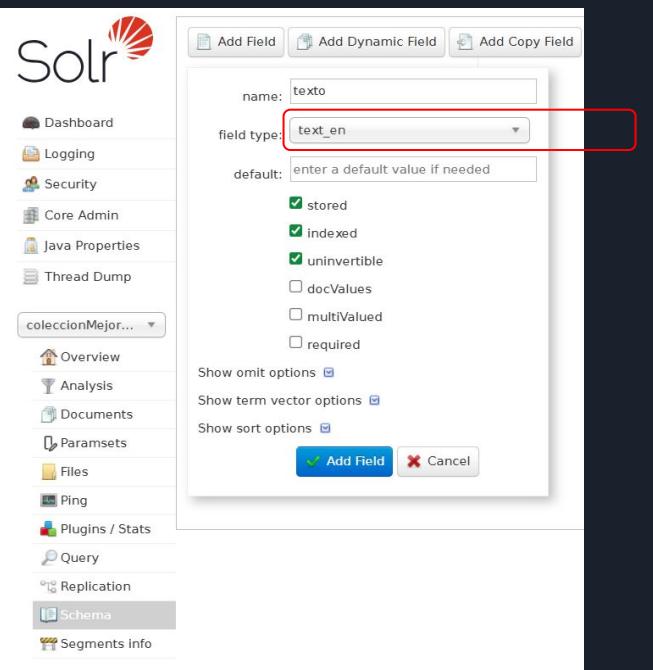


Mejoras del Motor de Búsqueda

Mejora 1

- Modificación del tipo de campo ‘texto’

Pantalla de Esquema

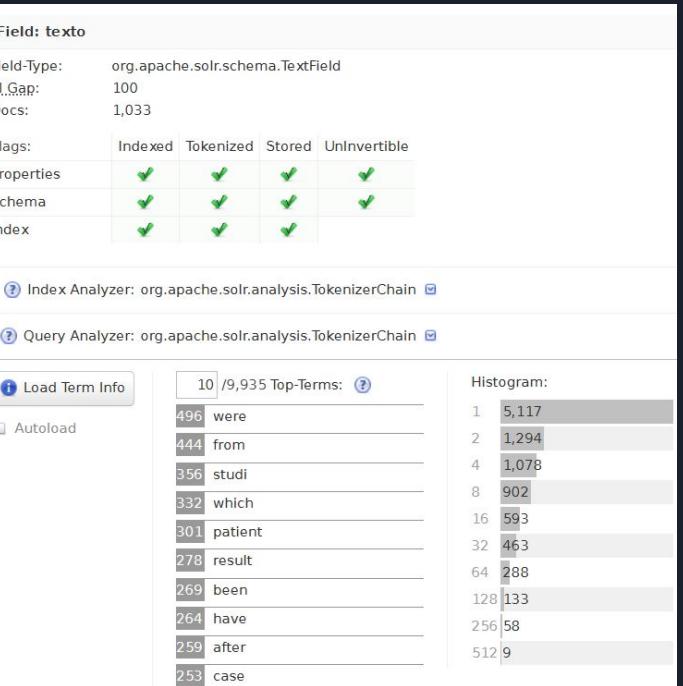


Mejoras del Motor de Búsqueda

Mejora 2

- Modificación de la lista de palabras vacías (Stopwords)

a	an	and
are	as	at
be	but	by
for	if	in
into	is	it
no	not	of
on	or	such
that	the	their
then	there	these
they	this	to
was	will	with



Mejoras del Motor de Búsqueda

Comparativa de Resultados

Recuperando **TODOS** los documentos

Queryid (Num): 30
Total number of documents over all queries
Retrieved: 28884
Relevant: 696
Rel_ret: 654
Interpolated Recall - Precision Averages:
at 0.00 0.9636
at 0.10 0.8019
at 0.20 0.7106
at 0.30 0.6639
at 0.40 0.5721
at 0.50 0.4668
at 0.60 0.4132
at 0.70 0.3405
at 0.80 0.2790
at 0.90 0.1798
at 1.00 0.0800
Average precision (non-interpolated) over all rel docs
0.4743
Precision:
At 5 docs: 0.7200
At 10 docs: 0.5967
At 15 docs: 0.5311
At 20 docs: 0.4767
At 30 docs: 0.3878
At 100 docs: 0.1700
At 200 docs: 0.0930
At 500 docs: 0.0404
At 1000 docs: 0.0216
R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: **0.4735**

Antes de las mejoras

Queryid (Num): 30
Total number of documents over all queries
Retrieved: 13506
Relevant: 696
Rel_ret: 629
Interpolated Recall - Precision Averages:
at 0.00 0.9515
at 0.10 0.8375
at 0.20 0.7773
at 0.30 0.7232
at 0.40 0.6089
at 0.50 0.5248
at 0.60 0.4438
at 0.70 0.3701
at 0.80 0.3024
at 0.90 0.2062
at 1.00 0.0736
Average precision (non-interpolated) over all rel docs
0.5175
Precision:
At 5 docs: 0.7200
At 10 docs: 0.6133
At 15 docs: 0.5733
At 20 docs: 0.5333
At 30 docs: 0.4167
At 100 docs: 0.1793
At 200 docs: 0.0987
At 500 docs: 0.0418
At 1000 docs: 0.0210
R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: **0.4940**

Después de las mejoras



Mejoras del Motor de Búsqueda

Comparativa de Resultados

Recuperando **30** documentos

Queryid (Num): 30
Total number of documents over all queries

Retrieved: 877
Relevant: 696
Rel_ret: 368

Interpolated Recall - Precision Averages:

at 0.00	0.9652
at 0.10	0.8035
at 0.20	0.7020
at 0.30	0.6353
at 0.40	0.5181
at 0.50	0.4051
at 0.60	0.2901
at 0.70	0.2205
at 0.80	0.1604
at 0.90	0.0446
at 1.00	0.0000

Average precision (non-interpolated) over all rel docs

0.4114

Precision:

At 5 docs:	0.7200
At 10 docs:	0.5967
At 15 docs:	0.5400
At 20 docs:	0.4867
At 30 docs:	0.4089
At 100 docs:	0.1227
At 200 docs:	0.0613
At 500 docs:	0.0245
At 1000 docs:	0.0123

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.4686

Antes de las mejoras

Queryid (Num): 30
Total number of documents over all queries

Retrieved: 900
Relevant: 696
Rel_ret: 384

Interpolated Recall - Precision Averages:

at 0.00	0.9515
at 0.10	0.8364
at 0.20	0.7611
at 0.30	0.6931
at 0.40	0.5824
at 0.50	0.4667
at 0.60	0.2933
at 0.70	0.2150
at 0.80	0.1603
at 0.90	0.0895
at 1.00	0.0000

Average precision (non-interpolated) over all rel docs

0.4439

Precision:

At 5 docs:	0.7200
At 10 docs:	0.6133
At 15 docs:	0.5689
At 20 docs:	0.5317
At 30 docs:	0.4267
At 100 docs:	0.1280
At 200 docs:	0.0640
At 500 docs:	0.0256
At 1000 docs:	0.0128

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.4892

Después de las mejoras



Ejecución

Iniciamos Apache Solr



```
C:\Apache Solr\solr-9.9.0-slim>bin\solr.cmd start
"Solr will start in SolrCloud mode by default in version 10, and you will have to provide --user-managed if you want to stay on the user-managed (aka. standalone) mode."
Java 21 detected. Enabled workaround for SOLR-16463
Java 21 detected. Incubating Panama Vector APIs have been enabled
Java HotSpot(TM) 64-Bit Server VM warning: JVM cannot use large page memory because it does not have enough privilege to lock pages in memory.
WARNING: A command line option has enabled the Security Manager
WARNING: The Security Manager is deprecated and will be removed in a future release
Waiting up to 180 seconds to see Solr running on port 8983
Started Solr server on port 8983. Happy searching!
```



LM Studio

Ejecución

Abrimos LM Studio

The screenshot shows the LM Studio application window. At the top, there's a toolbar with icons for file operations, a status indicator showing "Status: Running" with a toggle switch, and a dropdown menu "Select a model to load (Ctrl + L)". To the right, it displays "Reachable at: http://127.0.0.1:1234". On the left, there's a sidebar with icons for file, folder, and search. The main area has a message "No models loaded". Below that, a section titled "Supported endpoints (OpenAI-like)" lists various API endpoints with descriptions and links. At the bottom, there's a "Developer Logs" panel showing a list of log entries from December 13, 2025, at 20:41:59. The logs include messages about the server starting, supported endpoints, and just-in-time model loading.

```
2025-12-13 20:41:59 [INFO] [LM STUDIO SERVER] Success! HTTP server listening on port 1234
2025-12-13 20:41:59 [INFO]
2025-12-13 20:41:59 [INFO] [LM STUDIO SERVER] Supported endpoints:
2025-12-13 20:41:59 [INFO] [LM STUDIO SERVER] -> GET http://localhost:1234/v1/models
2025-12-13 20:41:59 [INFO] [LM STUDIO SERVER] -> POST http://localhost:1234/v1/responses
2025-12-13 20:41:59 [INFO] [LM STUDIO SERVER] -> POST http://localhost:1234/v1/chat/completions
2025-12-13 20:41:59 [INFO] [LM STUDIO SERVER] -> POST http://localhost:1234/v1/completions
2025-12-13 20:41:59 [INFO] [LM STUDIO SERVER] -> POST http://localhost:1234/v1/embeddings
2025-12-13 20:41:59 [INFO] [LM STUDIO SERVER] Logs are saved into C:\Users\Antonio Abad\lmstudio\server-logs
2025-12-13 20:41:59 [INFO] Server started
2025-12-13 20:41:59 [INFO] Just-in-time model loading active.
```



Ejecución

Abrimos NetBeans y ejecutamos el programa

The screenshot displays a Java Swing application window titled "Buscador Solr". At the top, there is a toolbar with a user icon and the text "Antonio Abad Hernandez Galvez -> Proyecto MB (Buscador Solr)". Below the toolbar, the main title bar reads "Buscador Solr". The interface includes a "Colección:" dropdown set to "colecciónMejorada" and a "tipo:" dropdown set to "texto". To the right of these dropdowns is a large search input field and a "Buscar" button. Below the search controls, there are two large text areas. The first text area is labeled "Respuesta generada por IA:" and the second is labeled "Respuesta:". Both text areas are currently empty.



Ejecución

Realizamos una búsqueda por texto

Antonio Abad Hernandez Galvez -> Proyecto MB (Buscador Solr)

Buscador Solr

Colección: colecciónMejorada texto ▾ eye

Buscar

Respuesta generada por IA:

Aquí hay información del contexto relacionada con los ojos:

- * **Documento 212:** Describe experimentos sobre el papel del humor acuoso y la retina en la regeneración del cristalino en los nuevos.
- * **Documento 730:** Trata las lesiones oculares oncocercales en monos.
- * **Documento 211:** Examina las reacciones de los tejidos del ojo y el riñón a los anticuerpos antiuveales heterólogos.
- * **Documento 213:** Se centra en la correlación entre las células de masto y el contenido de histamina del ojo en ganado.
- * **Documento 15:** Describe el desarrollo del cristalino, incluyendo experimentos in vitro e in vivo con células epiteliales del cristalino de pollos.
- * **Documento 512:** Presenta estudios posteriores de la epitelio del cristalino por autoradiografía en ojos de ratón normales y radiados con rayos X.
- * **Documento 104:** Analiza las respuestas evocadas en trastornos visuales.
- * **Documento 353:** Describe las manifestaciones oftálmicas de lesiones cerebrales bilaterales no occitales.
- * **Documento 623:** Presenta un resumen de los informes publicados sobre la histidinemia, incluyendo tres niños con esta condición.

Respuesta:

Se han recuperado un total de 16 documentos

Documento 212:

experiments dealing with the role played by the aqueous humor and retina in lens regeneration of adult newts . 1. these three groups of experiments involve approximately 140 eyes of adult newts, *triturus v. viridescens* . they were devised to examine what, if any, role the aqueous humor plays during lens regeneration from the dorsal iris . 2. many daily injections of aqueous humor from normal eyes were made in lentectomized eyes for as long as 96 days in some cases . as controls some lensless eyes were daily injected with holzfreter's solution . in others aqueous humor was merely withdrawn . 3) procedures for the injection experiments are difficult to control . however, the most successful cases showed varying degrees of inhibition and retardation of lens regeneration . 4. pairs of eyes were united at large adjacent wound openings to provide a common reservoir of aqueous humor bathing both lenses and dorsal irises . in some cases the eyes were placed on the side of the body . in others more successful unions were made by fusing a transplanted eye to the right eye of a host . 5. approximately three months after operation one of two large lens regenerates in a pair of perfectly fused eyes was removed . six weeks later a new large lens regenerate reappeared in most of the lentectomized units in the presence of the intact lens of the other

Ejecución

Respuesta LM Studio

```
[google/gemma-3n-e4b] Model generated tool calls: []
[google/gemma-3n-e4b] Generated prediction: {
  "id": "chatmpl-vrtjhef12qt4kxsf8fdi3",
  "object": "chat.completion",
  "created": 1765655262,
  "model": "google/gemma-3n-e4b",
  "choices": [
    {
      "index": 0,
      "message": {
        "role": "assistant",
        "content": "Aquí hay información del contexto relacionada con los ojos:\n\n**Documento 212:** Describe experimentos sobre el papel del humor acusoso y la retina en la regeneración del cristalino en los nuevos.\n**Documento 730:** Trata las lesiones oculares oncocercales en monos.\n**Documento 211:** Examina las reacciones de los tejidos del ojo y el riñón a los anticuerpos antiuveales heterólogos.\n**Documento 213:** Se centra en la correlación entre las células de masto y el contenido de histamina del ojo en ganado.\n**Documento 15:** Describe el desarrollo del cristalino, incluyendo experimentos in vitro e in vivo con células epiteliales del cristalino de pollos.\n**Documento 512:** Presenta estudios posteriores de la epitelio del cristalino por autoradiografía en ojos de ratón normales y radiados con rayos X.\n**Documento 104:** Analiza las respuestas evocadas en trastornos visuales.\n**Documento 353:** Describe las manifestaciones oftálmicas de lesiones cerebrales bilaterales no occipitales.\n**Documento 623:** Presenta un resumen de los informes publicados sobre la histidinemia, incluyendo tres niños con esta condición."
      },
      "tool_calls": []
    },
    "logprobs": null,
    "finish_reason": "stop"
  ]
},
"usage": {
  "prompt_tokens": 3814,
  "completion_tokens": 296,
  "total_tokens": 4110
},
"stats": {},
"system_fingerprint": "google/gemma-3n-e4b"
}
```



Ejecución

Realizamos una búsqueda sin sentido por texto

Antonio Abad Hernandez Galvez -> Proyecto MB (Buscador Solr)

Buscador Solr

Colección: colecciónMejorada

texto ▾ prueba

Buscar

NO se ha encontrado ningún documento.

Respuesta generada por IA:

Respuesta:



Ejecución

Realizamos una búsqueda por id

Antonio Abad Hernandez Galvez -> Proyecto MB (Buscador Solr)

Buscador Solr

Colección: colecciónMejorada id: 8

Buscar

Respuesta generada por IA:

No lo se aún, estoy trabajando en ello.

Respuesta:

Se han recuperado un total de 1 documentos

Documento 8:

essential fatty acids and acids with trans-configuration in the subcutaneous and visceral fat of the newborn . we made an investigation of the subcutaneous and visceral fat in the newborn . we estimated the contents of linoleic and linolenic acid and of acids with trans-configuration spectrophotometrically . we were able to show the penetration of these acids through the placental barrier . the essential fatty acid contents of fat in the newborn is low . in immature ones about 7-14 g, there is a rising trend.



¡Gracias!

Antonio Abad Hernández Gálvez

MOTORES DE BÚSQUEDA
Universidad De Huelva

