

Trabajo de Fin de Grado

perteneciente al alumno de Desarrollo de Aplicaciones Web (DAW)
Hugo Ruiz Sánchez, tutorizado por Julio Quiroga Alcalá, y
entregado el día 14 de junio de 2024.



Libre diccionario de la lengua española
Versión: 0.9.0

El documento carece de bibliografía, así como de referencias al código javascript de la interfaz gráfica; ya que el alumno se ha visto gravemente limitado por el tiempo de proyecto, debatiéndose entre completar la parte funcional o la documentación, ha primado la calidad del software sobre la calidad de la documentación.



Resumen o *abstract* del documento (Español)

El *Libre diccionario de la lengua española* (LDLE) es una iniciativa que aspira al desarrollo de recursos libres capaces de substituir efectivamente las funciones lexicográficas proporcionadas por las soluciones privadas facilitadas, muy especialmente, por el sitio dle.rae.es, respaldado por la Real Academia Española (RAE). Dichas funciones comprenden la ejecución de consultas por palabras – sea cual fuere su composición relativa al contexto –, contenido de las definiciones, anagramas, comienzo o término de las mismas.

Para la consecución de este objeto, presente proyecto ha logrado construir una base de datos en formato *JSON*, diseñada a propósito para ser empleada dentro de un programa que integre las funciones antedichas. Por añadidura a la anterior contribución, se ha desarrollado una aplicación web orientada al uso fuera de línea u *offline*, asimismo exportada en formato ejecutable nativo para los sistemas operativos *Microsoft Windows* (7 o posteriores), *Linux* o *GNU/Linux*.

Abstract of the document (English)

The Free (Libre) dictionary of the Spanish language (LDLE) is an initiative that aims at developing free resources capable of effectively replacing the lexicographic functions provided by private solutions, especially those provided by the site dle.rae.es, supported by the Real Academia Española (RAE). These functions include the execution of queries by words - whatever their composition relative to the context -, content of definitions, anagrams, beginning or end of words.

In order to achieve this goal, this project has succeeded in building a database in JSON format, purposely designed to be used within a program that integrates the above functions. In addition to the previous contribution, a web application has been developed for offline use, also exported in native executable format for Microsoft Windows (7 or later), Linux or GNU/Linux operating systems.

Índice

Sumario

Resumen o <i>abstract</i> del documento (Español).....	2
Abstract of the document (English).....	2
Índice.....	3
Justificación y antecedentes.....	3
Introducción.....	3
Objetivos.....	4
Estimación de costes del proyecto.....	5
Desarrollo del proyecto.....	5
Tecnología utilizada.....	5
Planificación.....	5
Especificación de requisitos.....	6
Diseño de la aplicación.....	6
Implementación y documentación.....	8
Plan de pruebas.....	11
Despliegue y mantenimiento.....	11
Conclusiones, mejoras futuras.....	12
Bibliografía y referencias.....	12
Anexos.....	12

Justificación y antecedentes

La Real Academia Española constituye una institución cultural privada, refrendada con oficialidad en calidad de autoridad reguladora de la lengua española. La naturaleza *mixta* sobre la que esta organización se erige deriva en las desventajas propias de las entidades privadas, en suma asimismo con los perjuicios sobre los contribuyentes que ocasionan las entidades públicas; esto es: la academia de la lengua, si bien financiada mediante el dinero de los contribuyentes españoles, aún se reserva el derecho con motivo de su carácter privado, de hacer públicos y abiertos los recursos de los que está provista, en beneficio de académicos, desarrolladores, o individuos cualesquiera interesados en el libre acceso a las bases de datos recopiladas por la RAE.

Introducción

La lengua es un objeto de imponderable valía, siendo que de ella pende el más elemental medio a través del cual toda información, pensamiento, instrucción, idea u operación puede trabarse. Las implicaciones de su ignorancia son profundas y no escapan a ningún campo o disciplina; inclusive en el que en el que compete a este proyecto: el desarrollo.

Trabajo de fin de grado, Libre diccionario de la lengua española. Hugo Ruiz Sánchez.

Para muestra de ello, los conocidos programas de denominada *inteligencia artificial*, especialmente aquellos clasificados como *chatbot*, en los que la lengua adquiere un eminente protagonismo, en tanto que es la base a través de la cual el cliente se comunica con el servicio. Estas gigantescas herramientas adquieren en buena medida un correcto desenvolvimiento en el habla gracias al análisis de cuantiosos volúmenes de bases de datos, antesala de los conocidos modelos de lenguaje en los que operan los conocidos servicios *ChatGPT* de *OpenAi*, *Llama* de *Meta* o *Gemini* de *Google*.

De consiguiente, la adecuación de la lengua sobre los medios informáticos precisos para su consulta tienen una relación directa con los avances tecnológicos han acontecido recientemente. Si estos recursos están cerrados, o no hay personas dispuestas para llevarlos a cabo, haremos de nuestra lengua una materia rezagada para las inteligencias; todavía hoy, al remitirnos en español a los populares programas de conversación, podrá apercibirse un sustancial decrecimiento en la comprensión con respecto al idioma inglés; la causa, presumiblemente, radica en la falta de recursos de que están dispuestos los desarrolladores.

Objetivos

El objetivo primordial de *L·DLE* es construir una base de datos comportada por la totalidad de las palabras existentes en la lengua castellana, orientada su empleo por medio de clientes de consulta; para ese fin, estará cuidadosamente provista de los nodos efectivos para llevar a cabo cualquier tipo de petición en relación a un lema, y recibir una respuesta con independencia de su contexto o forma.


Por añadidura, el éxito de esta empresa se valorará con la creación de un cliente multiplataforma – basado en la web –, capaz de, a partir de una entrada, recibir resultados semejantes a los esperados por los medios de la Real Academia Española en su sitio oficial, dle.rae.es.

Diccionario de la lengua española

Edición del Tricentenario

Actualización 2023

Consulta posible gracias al compromiso con la cultura de la

 Fundación "la Caixa"

por palabras

▼

Escriba aquí la palabra

Consultar

Diccionario de la lengua española

El *Diccionario de la lengua española* es la obra lexicográfica académica por excelencia.

El repertorio empieza en 1780, con la aparición —en un solo tomo para facilitar su consulta— de una nueva versión, ya sin citas de autores, del primer diccionario de la institución, el llamado *Diccionario de autoridades* (1726-1739). El de 1780 fue, por tanto, el precedente de la serie de diccionarios usuales que llega hasta hoy.

Desde entonces, se han publicado [veintitrés ediciones](#) de la obra, convertida, a través del tiempo, en el diccionario de referencia y consulta del español. La más reciente, la [23.^a](#), salió de imprenta en octubre de 2014.

El *Diccionario de la lengua española* es el resultado de la colaboración de [todas las academias](#), cuyo propósito es recoger el léxico general utilizado en España y en los países hispanicos. Se dirige, fundamentalmente, a hablantes cuya lengua materna es el español, quienes encontrarán en él recursos suficientes para descifrar textos escritos y orales.

Estimación de costes del proyecto

El desarrollo de software por lo general no requiere inversiones iniciales, más que para disponer de los medios de almacenaje adecuados, en este caso un, *pen-drive* de 5.99 euros.

Desarrollo del proyecto

Tecnología utilizada.

Las tecnologías para llevar a cabo este proyecto serán *libres* o de *código abierto*, condiciones gracias a las cuales podrá hacerse uso de las mismas sin repercutir sobre la nula estimación de costes asumida anteriormente: no habrá costo añadido, no en lo respectivo a las tecnologías, así tampoco en las herramientas para programarlas.

La creación de la base de datos se realizará sobre el lenguaje de programación *python*, utilizando PyDev de Eclipse para la programación y ejecución de los scripts, y VIM como visualizador preferido para comprobar la recolección de los datos, especificada en formato JSON. Respecto al uso de *Python*, se subraya el uso de importantes librerías instaladas con pip para tareas concretas tales como: serialización en formato binario o JSON, funciones lexicográficas especiales, exportación y tratamiento de archivos, etc...

El desarrollo de la GUI se realizará desde *VsCodium*, un fork *Open Source* del editor de código de Microsoft, *VsCode*. Esta será programada enteramente como una página estática basada en javascript, implementando la base de datos previamente creada. La ejecución de la web para realizar las pruebas pertinente sobre la misma se realizará sobre los navegadores *Chromium* y *Mozilla Firefox*.

La construcción del proyecto web para convertirlo a nativo se logrará mediante el framework de *Electron*, Gracias a este, el

Planificación

Dado que la creación de la base de datos figura dentro de los objetos primordiales de este proyecto, siendo además que esta es pieza imprescindible para el buen funcionamiento del cliente web, dos meses y medio de los tres que serán dedicados al proyecto estarán dedicados a la realización de los *scripts* en *python* que generen una base de datos de la mayor calidad posible.

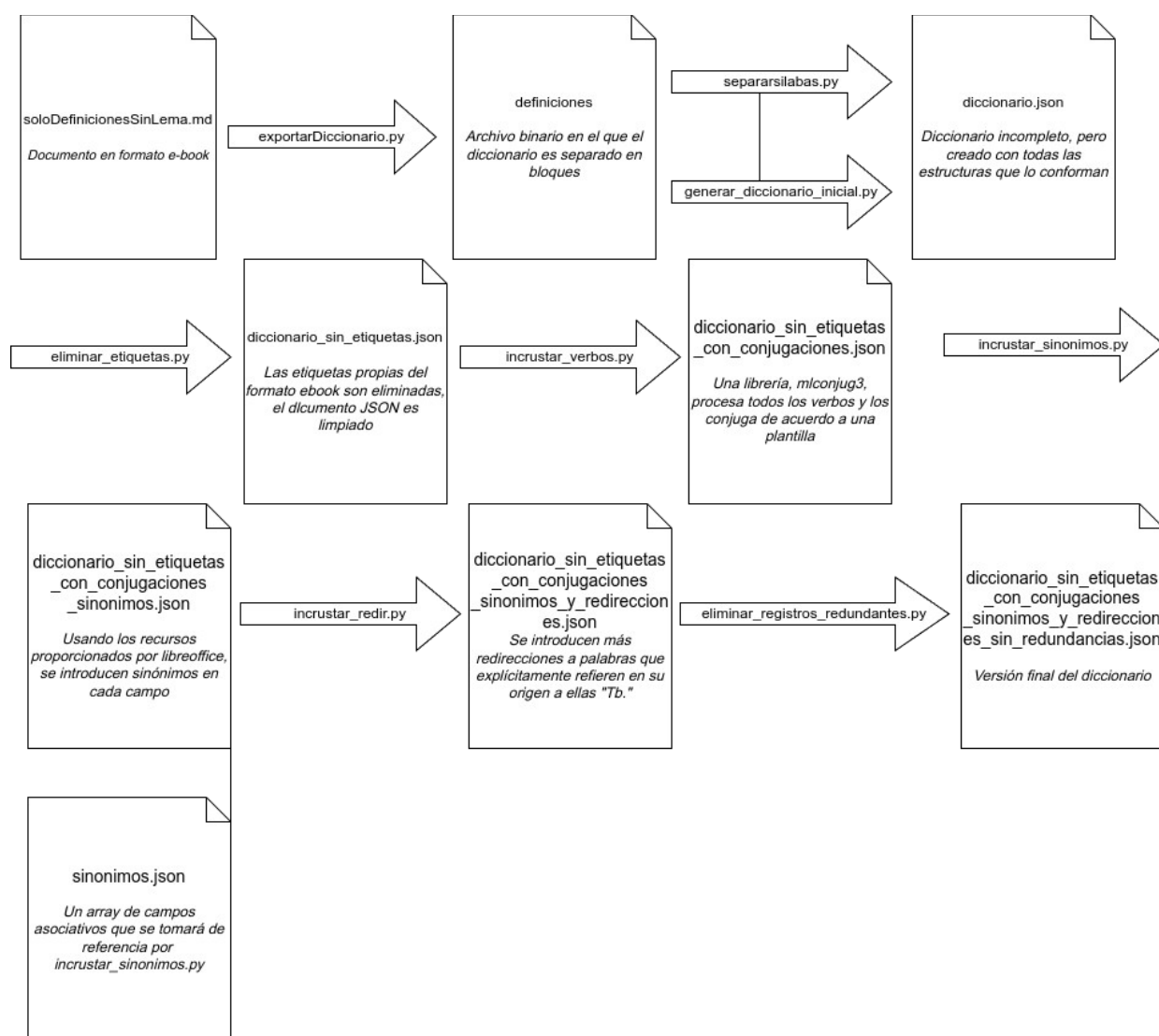
Especificación de requisitos

El uso de esta herramienta será compatible con cualquier sistema operativo, sea gracias a los archivos ejecutables que lo harán multiplataforma: paquetería *appimage* para sistemas operativos Linux de 32 y 64 bits, presentada en un formato específico que contiene todas las dependencias del proyecto indistintamente de la distribución en que se ejecute; para los sistemas operativos de Microsoft Windows, también se dispondrá de instaladores *.setup* ejecutables *.exe*, admisibles solo para versiones que aún cuenten con el soporte oficial brindado para *Chromium*.

Puesto que es, a todos los efectos, una aplicación web, cualquier otro sistema operativo no recogido podrá seguir accediendo al diccionario desde el navegador.

Diseño de la aplicación.

Creación del diccionario.



El contenido de la base de datos, sus redirecciones, apartados y vínculos pueden visualizarse mediante una herramienta realizada para facilitar el trabajo de resolución de errores y comprensión del sistema, *consultar diccionario.py*, que accede al último fichero generado.

```
Li.DRAE - Consulta a la base de datos
Cargando ficheros JSON . . .
. . . Ficheros JSON cargados correctamente.
<ldle> redingote
[
  {
    "palabra": "redingote",
    "origen": [
      "Del fr. *redingote*, y este del ingl. *riding-coat* 'traje para montar'"
    ],
    "sinonimos": [],
    "definicion": "m. Capote de poco vuelo y con mangas ajustadas.",
    "formas_compuestas": [],
    "envios": []
  }
]
<ldle> □
```

El diccionario dispone de un buscador. Podrá hacer sugerencias a medida que escribes palabras en él.





Los apartados de la barra lateral permiten realizar búsquedas según lo especificado: búsqueda exacta, empieza por, termina en, contiene, anagramas o búsqueda de palabras aleatoria.

L·DLE

exacta amigo

amigo, ga

amigo, ga

Del lat. *amīcus*
Sup. irreg. **amícsimo**; reg. **amiguísimo**

1. adj. Que tiene relación de amistad. U. t. c. s.
2. **amistoso** (perteneciente a la amistad).
3. Dicho de una cosa: Propia de la persona **amiga**.
4. Que gusta mucho de algo. *Es amigo de trasnochar*. U. t. c. s. *Los amigos del silencio disfrutarán en este bosque*.
5. poét. Dicho de una cosa: Propicia, benigna, grata. *El cielo dio su amiga sombra al peregrino*.
6. m. y f. U. para dirigirse a una persona, aunque no se conozca o no haya relación de amistad.
7. U. en cartas como tratamiento afectuoso, aunque no haya propiamente relación de amistad.
8. **amante** (persona que mantiene con otra una relación amorosa).
9. f. p. us. Maestra de escuela de niñas.
10. *And.* y *Méx.* Escuela de párvulos.

amigo, ga de lo ajeno

1. m. y f. coloq. Ladrón, estafador.

amigo, ga de pelillo , o amigo, ga de taza de vino

1. m. y f. coloqs. desus. **amigo** que lo es solamente por interés y conveniencia.

amigo, ga del asa


1. m. y f. coloq. desus. **amigo** íntimo.

Cada palabra puede disponer de *origen*, *definiciones*, *formas compuestas*, *sinónimos*, *expresiones*, *envíos*.

Trabajo de fin de grado, Libre diccionario de la lengua española. Hugo Ruiz Sánchez.

En ocasiones, podrá resultar que una palabra contenga conflictos no resueltos en la base de datos: se pedirá al usuario que los resuelva.

L·DLE



amigas

amigo, ga


Entradas que contienen la forma «amigas»:

amigar

amigo, ga

El diccionario también hace uso de un historial, lo que permite mantener un registro de todas las palabras buscadas.

L·DLE



sumatra

liderear

ocioso, sa

amigas

amigo, ga

La palabra «sumatra» no está en el diccionario. Las entradas que se muestran a continuación podrían estar relacionadas:

sumaca

sumar

sumir

sumara

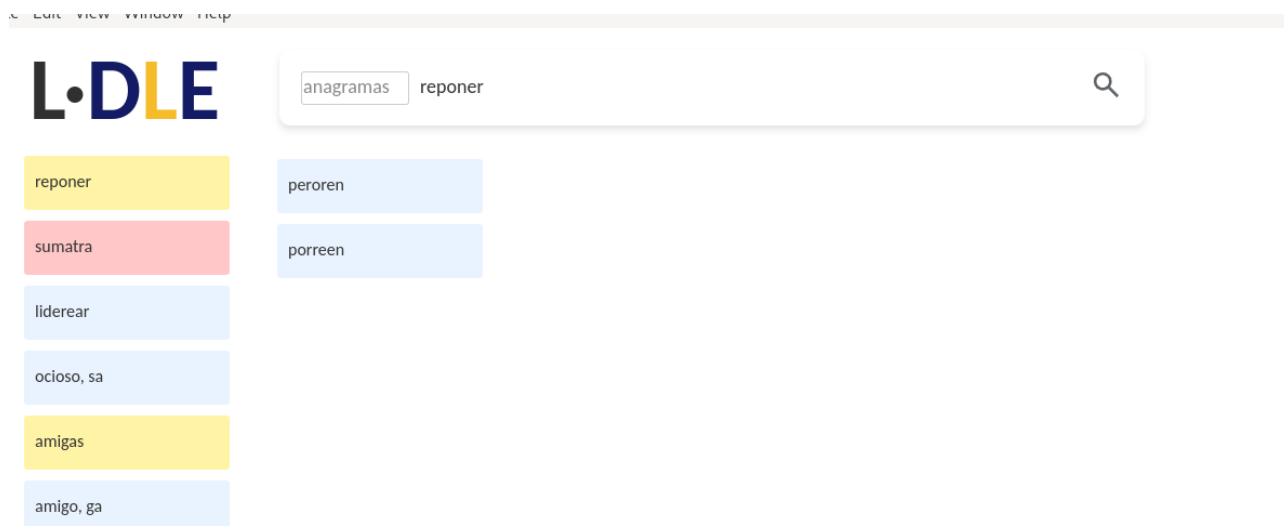
sumaba

sumad

suman

sumas

Si la palabra es incorrecta, el corrector buscará similitudes existentes con otras palabras que constan en el diccionario.



La búsqueda de anagramas, terminaciones o comienzos también deriva al mismo estado que cuando existen conflictos en la base de datos.

Plan de pruebas.

Las pruebas sobre la base de datos se realizan mediante algunas secuencias de código comentadas e incrustadas en el fichero de *verificar_integridad.py*, que toma como referencia el fichero *0_palabras_todas.txt*, en el que están contenidas todas las palabras de la lengua castellana. Si uno de las palabras no existe, se ejecutan una serie de secuencias sobre la palabra solicitada para verificarlo; si después de ese procedimiento la base de datos no da respuesta, ni se ha verificado que tal palabra esté vinculada a

Despliegue y mantenimiento.

La creación del diccionario tan solo precisa de ejecutar, en el orden expuesto anteriormente, cada uno de los scripts generando sucesivamente cada uno de los archivos *json* resultantes hasta su versión definitiva, la que tomará de referencia el ejecutable *consultar_diccionario.py* y la que se recomienda usar para la GUI.

El despliegue del aplicativo es tan sencillo como ejecutar los ficheros portables de ejecución si se trata de usarlos nativamente en el equipo, o subirlos a un servidor como ficheros html convencionales, siendo que todas las operaciones se desarrollarán del lado del cliente.

Trabajo de fin de grado, Libre diccionario de la lengua española. Hugo Ruiz Sánchez.

La compilación mediante electron está automatizada gracias a unos scripts presentados en *package.json*, sobre los que debe realizarse simplemente `npm run nombre_script`

Conclusiones, mejoras futuras.

A lo largo del código se encuentran sugerencias de mejora. La más imperiosa de todas ellas es la implementación de las palabras restantes, pertenecientes a nuevas reediciones de la RAE, y a las que se prevee acceder mediante web scrapping; este proceso ya ha sido esbozado en un script, denominado *incrustar_palabras_restantes.py*.

Bibliografía y referencias

Anexos