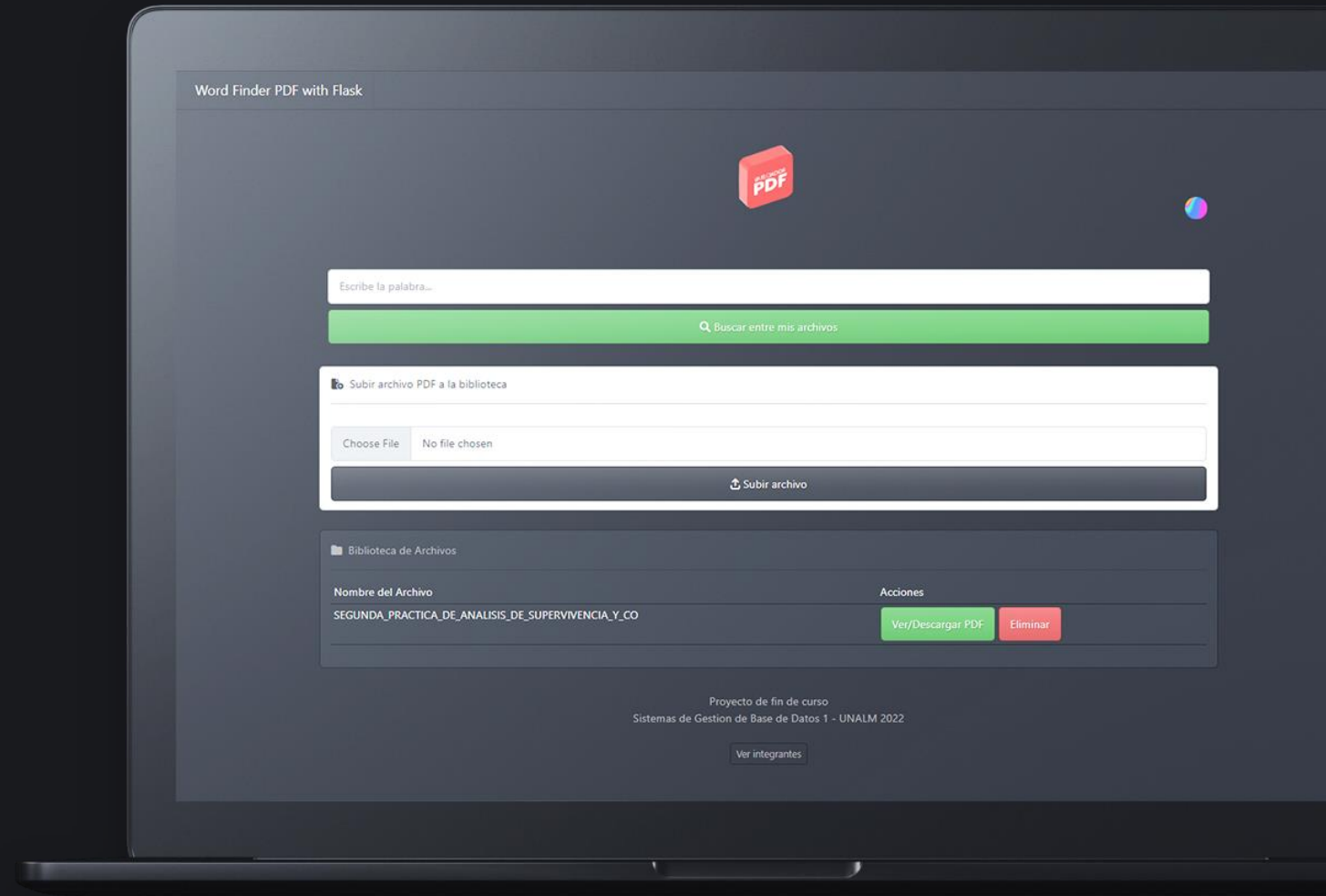


BUSCADOR PALABRAS EN PDF

Integrantes:

- Gomez Mendez Yan Franco
- Gutarra Seguil Gabriela Gahia
- Lopez Ore, Jimmy Arthur
- Morales Morales Flavio Oscar
- Roca Llamocca, Roger Ricardo



Guía de Uso

Buscar palabra

Escribe la palabra que quieras ubicar dentro de todos tus archivos PDF subidos

1

Escribe la palabra...

🔍 Buscar entre mis archivos

Subir archivos

Elije tus archivos PDF para subirlo a la biblioteca

2

📎 Subir archivo PDF a la biblioteca

Choose File

No file chosen

📤 Subir archivo

Lista de archivos

Podrás visualizar tus archivos PDF o eliminarlo de la biblioteca

3

📁 Biblioteca de Archivos

Nombre del Archivo

Acciones

SEGUNDA_PRACTICA_DE_ANALISIS_DE_SUPERVIVENCIA_Y_CO

Ver/Descargar PDF

Eliminar

Resultado al buscar palabra

🔍 Resultados

Visualización del nombre de los PDF's que incluyen el texto buscado además de todo el contenido de texto extraído del archivo

[← Regresar al inicio](#)

Palabra buscada: 'supervivencia'

📁 Resultados de Búsqueda...

Nombre del Archivo	Contenido	Acciones
SEGUNDA_PRACTICA_DE_ANALISIS_DE_SUPERVIVENCIA_Y_CO	segunda practica de analisis de supervivencia y confiabilidad estos datos del registro de tumores de connecticut presentan las cifras ajustadas a la edad de las incidencias de cancer de piel de melanoma por cada 100.000 personas en connecticut para los años de 1936 a 1972. la base de datos melanoma contiene un total de 205 observaciones de 6 variables: 1. no. codigo del paciente. 2. status. el estado de los pacientes y el final del estudio (1=muere por melanoma maligno, 2=vive al final del estudio, 3=muere por otras causas). 3. days. tiempo de estudio despues de la operacion con melanoma maligno. 4. ulc. superficie del melanoma vista por un microscopio si muestra ulceracion (1=presente, 2=ausencia). 5. thick. el grosor del tumor. 6. sex. genero de los pacientes (1=femenino, 2=masculino). 1) realice las pruebas logrank, wilcoxon y tarone-ware para la igualdad de las dos funciones de supervivencia estimadas para el grupo masculino y femenino. 2) empleando el software r para grupo femenino, masculino y ambos encuentre: a) la funcion de densidad exponencial o weibull (grafico) b) la funcion de distribucion acumulada (grafico) c) la funcion de sobrevivencia (grafico) d) la funcion de riesgo (grafico) e) funcion de riesgo acumulado (grafico) f) interpretar cada grafica 3) compare la grafica de funcion de supervivencia parametrica y no parametrica para grupo femenino, masculino y ambos. 4) encuentre el mejor modelo de riesgos proporcionales; interprete los coeficientes	<div>Ver/Descargar PDF</div> <div>Eliminar</div>

Recursos utilizados



Python

+



Framework Flask



HTML



CSS

+



Bootstrap



Java Script



MySQL



Front-end



HTML

+3



CSS

Estilos



JavaScript

```
const btnDelete= document.querySelectorAll('.btn-delete');
if(btnDelete) {
  const btnArray = Array.from(btnDelete);
  btnArray.forEach((btn) => {
    btn.addEventListener('click', (e) => {
      if(!confirm('¿Estas seguro de eliminar este archivo PDF?')){
        e.preventDefault();
      }
    });
  });
}
```

Botón

Back-end



Python

```
# Librerias
import os
from os import remove
import datetime
import pandas as pd
from flask import Flask, flash, render_template, request, redirect, url_for
from flask_mysql import MySQL
from werkzeug.utils import secure_filename
from werkzeug.datastructures import FileStorage
```

```
# Conexión con MySQL
app.config['MYSQL_HOST'] = '50.31.188.40'
app.config['MYSQL_USER'] = 'yvalpqfp_roger_test'
app.config['MYSQL_PASSWORD'] = 'nju30xnv5xib'
app.config['MYSQL_DB'] = 'yvalpqfp_sgbd1_test'
mysql = MySQL(app)
```

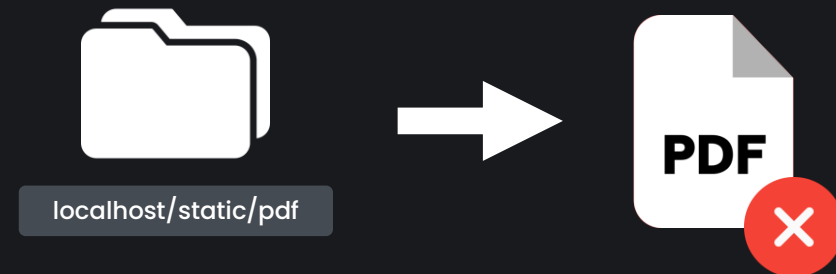
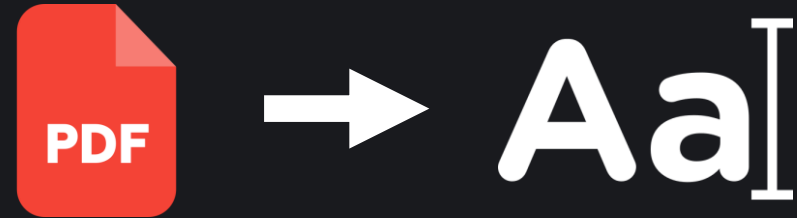


Back-end



```
# Función para convertir pdf a texto
def convert_pdf_to_text(doc_pdf):
    from tika import parser
    file = doc_pdf
    file_data = parser.from_file(file)
    text = normalize(file_data['content'])
    return text.encode('utf-8') # Convirtiendo a UTF-8
```

```
# Ruta para eliminar un pdf de la base de datos
@app.route('/delete/<string:id>')
def delete_contact(id):
    cur = mysql.connection.cursor()
    cur.execute('DELETE FROM archivos_pdf WHERE id_archivo={0}'.format(id))
    print(id)
    mysql.connection.commit()
    remove("static/pdf/file-pdf-{0}.pdf".format(id))
    flash('Archivo Removido Satisfactoriamente')
    return redirect(url_for('Index'))
```





```
@app.route('/search-results')
def searchresults():
    return render_template('search-results.html')
```

Gestionar redirecciones



```
@app.route('/', methods=['POST'])
def my_form_post_search():
    text = request.form['text_search_web']
    cur = mysql.connection.cursor()
    cur.execute("SELECT * from archivos_pdf WHERE contenido LIKE '%" + str(normalize(text)) + "%'")
    data_search = cur.fetchall()
    return render_template('search-results.html', data2=data_search, textsearch="Palabra buscada: '" + normalize(text) + "'")
```

Escribe la palabra...

🔍 Buscar entre mis archivos



Table 'archivos_pdf'



MySQL

Server: localhost:3306 » Database: yvalpqfp_sgbd1_test » Table: archivos_pdf

Browse Structure SQL Search Insert Export Import Operations Triggers

✓ Showing rows 0 - 1 (2 total, Query took 0.0003 seconds.)

```
SELECT * FROM `archivos_pdf`
```

☐ Profiling [Edit inline] [Edit] [Expla]

☐ Show all | Number of rows: 25 | Filter rows: Search this table | Sort by key: None

+ Options

	id_archivo	nombre_archivo	fecha_hora_subida	ruta_pdf	ruta_text	contenido
<input type="checkbox"/> Edit Copy Delete	48	SEGUNDA_PRACTICA_DE_ANALISIS_DE_SUPERVIVENCIA_Y_CO	2022-08-08 19:29:23	static/pdf/file-pdf-48.pdf	static/pdf/file-text-48.txt	segunda practica de analisis de superv...
<input type="checkbox"/> Edit Copy Delete	49	2022-06-07T00_00_000001288163.pdf	2022-08-10 16:18:14	static/pdf/file-pdf-49.pdf	static/bt/file-text-49.txt	report s/ 34.10 22/06/2022 * consumo ...

↑ ☐ Check all | With selected: Edit Copy Delete Export

☐ Show all | Number of rows: 25 | Filter rows: Search this table | Sort by key: None

Query results operations

Print Copy to clipboard Export Display chart Create view

Base de datos



```
# Ruta para subir los documentos PDF
@app.route("/upload", methods=['POST'])
def uploader():
    if request.method == 'POST':
        # Obtenemos el archivo del input "archivo"
        f = request.files['archivo']
        filename = secure_filename(f.filename)
        print(filename)

        # Definir el id para el nuevo archivo subido
        cur = mysql.connection.cursor()
        cur.execute('SELECT MAX(id_archivo) FROM archivos_pdf;')
        id_archiv = cur.fetchall()[0][0] + 1
        print(id_archiv)

        # Guardamos el archivo en el directorio "static/pdf/"
        f.save(os.path.join(app.config['UPLOAD_FOLDER'], "file-pdf-" + str(id_archiv) + ".pdf"))

        txt_temp = convert_pdf_to_text('static/pdf/' + "file-pdf-" + str(id_archiv) + ".pdf").lower()

        # Definir la fecha y hora de subida
        fecha_temp = datetime.datetime.now()

        # Preparando variables a almacenar en base de datos
        nombre_archivo = filename
        ruta_pdf = "static/pdf/file-pdf-" + str(id_archiv) + ".pdf"

        # Conectando a una base de datos y insertando los valores nuevos
        cur = mysql.connection.cursor()
        cur.execute('INSERT INTO archivos_pdf (nombre_archivo, fecha_hora_subida, ruta_pdf , ruta_text, contenido) VALUES (%s, %s, %s, %s, %s)',
        (nombre_archivo, fecha_temp, ruta_pdf, '', txt_temp))
        mysql.connection.commit()
        txt_temp=''
        fecha_temp=''

        # Retornamos una respuesta satisfactoria
        flash('Archivo ' + filename + ' subido exitosamente')
        return redirect(url_for('Index'))
```



Guardar el archivo en local /static/pdf



Escribir nombre, fecha/hora registro, ruta en local y contenido del texto en la tabla 'archivos pdf'

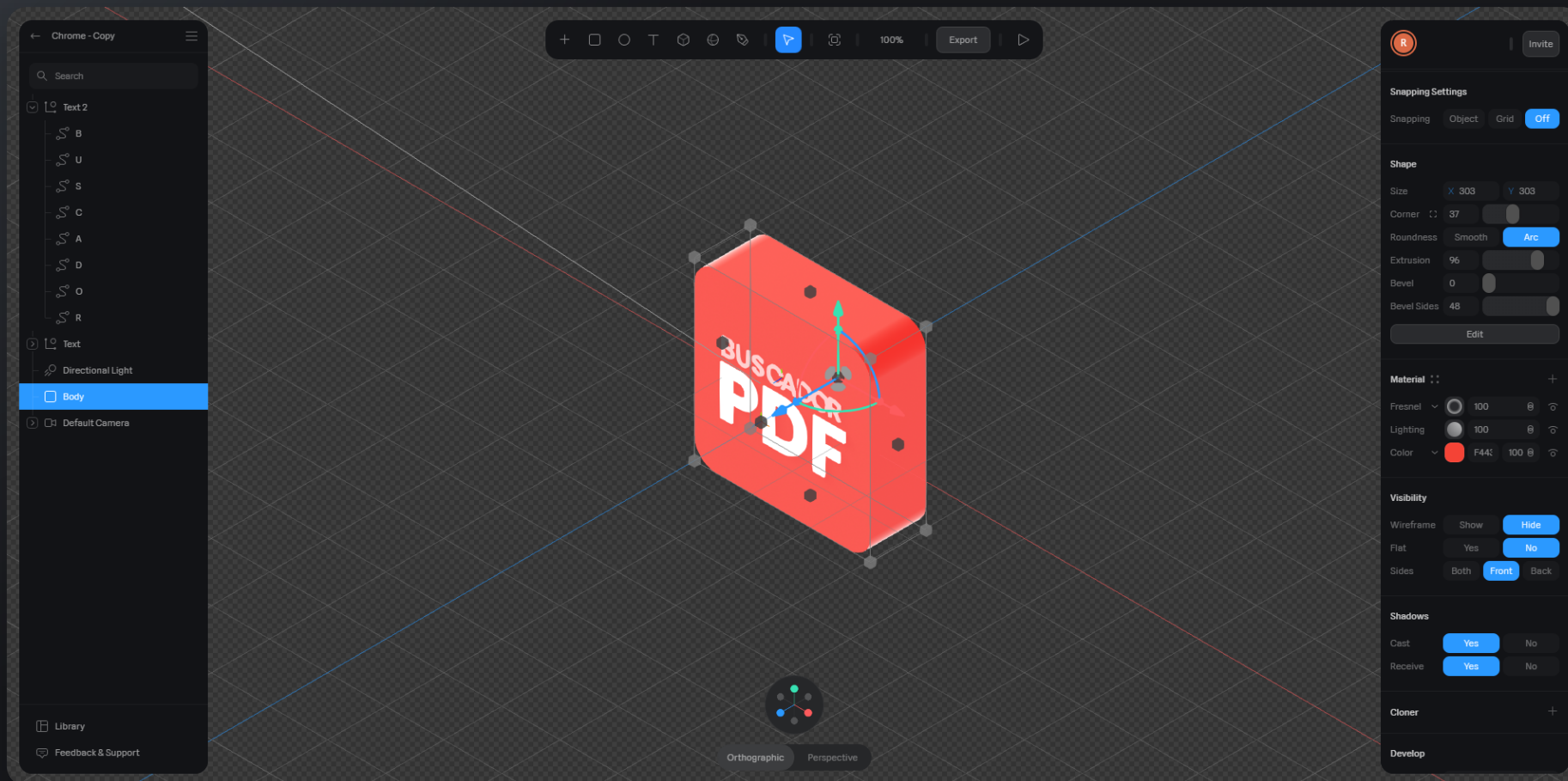


Recursos Adicionales



Spline

<https://spline.design>





Bootstrap

https://getbootstrap.com

 [Docs](#) [Examples](#) [Icons](#) [Themes](#) [Blog](#)

     v5.2 ▾


New in v5.2 CSS variables, responsive offcanvas, new utilities, and more!



Build fast, responsive sites with Bootstrap

Powerful, extensible, and feature-packed frontend toolkit. Build and customize with Sass, utilize prebuilt grid system and components, and bring projects to life with powerful JavaScript plugins.

```
$ npm i bootstrap@5.2.0
```

 [Read the docs](#)

Currently **v5.2.0** · [Download](#) · [v4.6.x docs](#) · [All releases](#)

<>

Get started any way you want

Framework para HTML, CSS y JavaScript

Recomendaciones



Alternativas a *Tika*

Al no ser la única librería en realizar la misma función de lectura de PDF's, hay otros como **PyPDF2**, **pdfPlumber**, **pdfMiner**, etc que pueden dar mejores resultados en ciertos archivos.

Mejorar apariencia de los resultados de búsqueda

Al buscar una palabra, da como resultado en una de las columnas todo el contenido de texto del archivo PDF, además que no se ubica fácilmente la palabra, sino haciendo un **ctrl + F**.

Implementación de usuario

Poder registrarte y ingresar con una cuenta propia que registre tus propios archivos de manera independiente.



Gracias por su atención

