Document-search-references-update

In document search the model answers based on collection of chunks of text extracted from the ingested files depending on the user’s question.

(Eg : Spendmate.pdf(4 pages) will be about 4 chunks of text)
As citations for each user question we can have the following options.
Question : What is spendmate ?
1. Chatbot Response /n +
   References : The textual sections that were used. (Will show a list of ‘n’ citations. where we fix the number n during the chatbot logic design.)

2. Chatbot Response /n +
   References : The textual sections that were used. +
   URL : to open/view the actual source file related to the picked up content. (whole pdf/file will be visible when clicked)

```
EngageMate© is an elegant system that solves all the problems associated with the engagement

lifecycle. Recently, after an unsatisfactory experience with another engagement tool, a top pharmaceutical company returned to
tracking their activities with emails, spreadsheets, and word documents. Cresen Solutions was then brought in to implement Enga
geMate©. After gathering business requirements around workflow requirements, business rules, and input forms EngageMate© was co
nfigured to meet all of the client’s policies and SOPs. EngageMate© was integrated with their customer

666 Exton Commons Exton PA 19341 (484) 879-426

master to pull service provider information into the system, ensuring accurate Customer(HCP or HCO) details while limiting manu
al data entry. Review workflows were created for each region within the company enabling a smooth process for the activity to b
e reviewed by multiple departments and reviewers across the globe. After the workflows are completed EngageMate© assures proper
closure and certification processes aligned with company policy. EngageMate© uses all the latest technologies and uses common o
pen integration methodologies to enable easy interfacing with other company business systems.

From adding service providers to filling out forms and submitting activities, business users have

enjoyed EngageMate’s© responsiveness and ease of use. Most users were able to pick up the navigation and use of the system afte
r one self-guided training session. Everyone, both “Submitters” and “Reviewers”, receive email notifications when an activity i
s ready for their action. The email provides an embedded hyperlink that takes users directly to the activity they need to take
action on. Additionally, upon logging into EngageMate©, users are presented with a simple landing page showing tiles that separ
ate activities into distinct categories allowing users to see a filtered list of activities to find any activity they are looki
ng for expeditiously. It also includes customizable dashboards that pull

Source: C:\Users\ANANTH~1\AppData\Local\Temp\tmp4x_q5b2i\apps\mm\jazz\6.Cresen_whitepapers\EngageMate Whitepaper.pdf
URL: https://cresendrugdata.blob.core.windows.net/cresendev/apps/mm/jazz/6.Cresen_whitepapers/EngageMate Whitepaper.pdf?se=2023
-10-30T12%3A24%3A41Z&sp=r&sv=2021-08-06&sr=b&sig=DUNFgbqO1%2BzfSeQnt5I6JMhN0Yf86w4kI1nP2TqC5yg%3D
```

the blob uri is not accessible anonymously/public. It is processed through a function which take the uri as input and through our access tokens it will generate a sas token

and subsequent sas url which will be active for one hour(which we can specify). These endpoints ensure privacy of the files. This final sas url can be ingested in an html file and viewed.

```
: from azure.storage.blob import generate_blob_sas, BlobSasPermissions
  from datetime import datetime, timedelta
  import urllib.parse

  ACCOUNT_NAME = account_name
  ACCOUNT_KEY = account_key
  CONTAINER_NAME = 'cresendev'

  def generate_sas_token(blob_name):
      # Generate a SAS token
      sas_token = generate_blob_sas(
          account_name=ACCOUNT_NAME,
          container_name=CONTAINER_NAME,
          blob_name=blob_name,
          account_key=ACCOUNT_KEY,
          permission=BlobSasPermissions(read=True),
          expiry=datetime.utcnow() + timedelta(hours=1)
      )

      blob_url_with_sas = f'https://{ACCOUNT_NAME}.blob.core.windows.net/{CONTAINER_NAME}/{blob_name}?{sas_token}'
      return blob_url_with_sas

  def generate_sas_for_url(url):
      # Parse the URL to get the blob name
      blob_name = url.split(f'https://{ACCOUNT_NAME}.blob.core.windows.net/{CONTAINER_NAME}/')[-1]

      # Decode the URL-encoded blob name
      blob_name = urllib.parse.unquote(blob_name)

      # Generate the SAS token
      sas_url = generate_sas_token(blob_name)
      return sas_url
```

index.html

File Edit View

```
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>PDF Viewer</title>
<style>
  #pdf-container {
    display: flex;
    flex-direction: column;
    align-items: center;
  }
  canvas {
    margin-bottom: 20px;
  }
</style>
</head>
<body>
<div id="pdf-container"></div>
<script src="https://cdnjs.cloudflare.com/ajax/libs/pdf.js/2.6.347/pdf.min.js"></script>
<script>
var url = 'https://cresendrugdata.blob.core.windows.net/cresendev/apps/mm/jazz/6.Cresen_whitepapers/SpendMate%20Witepaper.pdf?
se=2023-10-30T10%3A36%3A06Z&sp=r&sv=2021-08-06&sr=b&sig=76gcDWLJUdSWz4qRLo54MjZUUIzdbeY18XiJW3PAlAM%3D';

var pdfjsLib = window['pdfjs-dist/build/pdf'];
pdfjsLib.GlobalWorkerOptions.workerSrc = 'https://cdnjs.cloudflare.com/ajax/libs/pdf.js/2.6.347/pdf.worker.min.js';

pdfjsLib.getDocument(url).promise.then(function(pdf) {
  var pdfContainer = document.getElementById('pdf-container');

  function renderPage(page) {
    var scale = 1.5;
```

3. Chatbot Response /n +

References : The textual sections that were used. +

URL : to open/view the the split up pdf page only, (Here during ingestion we split the spendmate.pdf into 4 pdfs for each page and the its processed. In this way the viewing

of the file will pinpoint to the related section more precisely than showing all the pages)

And all of these can either be implemented within the same chatbox (depends on how the other elements of the page will adjust)

or the references can be opened up in a new tab through a hyperlink without exposing any of the links/paths.