

日々是Oracle APEX

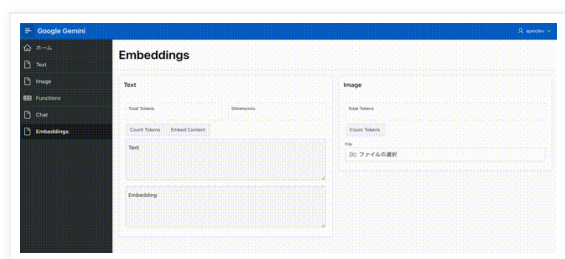
Oracle APEXを使った作業をしていて、気の付いたところを忘れないようにメモをとります。

2024年1月9日 火曜日

Google GeminiのCount TokensとEmbed Contentを呼び出す

Google Geminiを呼び出すアプリケーションにトークン数のカウントとベクトル埋め込み（embeddings）の生成を行うページを追加しました。

以下のように動作します。



文章またはファイルのデータを送信して、トークン数またはベクトル埋め込みをレスポンスとして受け取ります。

ボタンCOUNT_TOKENSのクリックで、ページ・アイテムP7_TEXTの文章をリクエストに含め、Google GeminiのcountTokens APIを呼び出して返されたトークン数が、P7_TOTAL_TOKENSに設定されます。

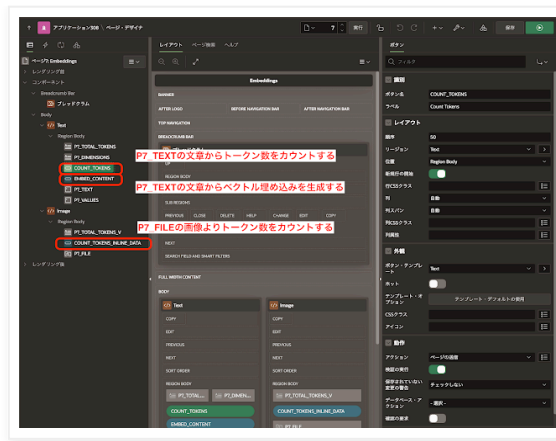
ボタンEMBED_CONTENTのクリックで、ページ・アイテムP7_TEXTの文章をリクエストに含め、Google GeminiのembedContent APIを呼び出して返されたベクトル埋め込みがP7_VALUESに、返されたベクトル埋め込みの次元数がP7_DIMENSIONSに設定されます。

ボタンCOUNT_TOKENS_INLINE_DATAのクリックで、ページ・アイテムP7_FILEの画像ファイルをリクエストに含め、Google GeminiのcountTokens APIを呼び出して返されたトークン数が、P7_TOTAL_TOKENS_Vに設定されます。

Google Geminiのベクトル埋め込みを生成するためのモデルmodels/embedding-001の入力はテキストのみです。画像からベクトル埋め込みを生成できるモデルは、以下のGeminiモデルの一覧には見当たりませんでした。

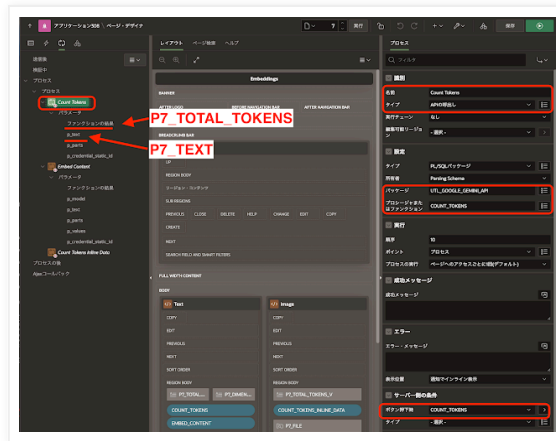
<https://ai.google.dev/models/gemini>

（この記事の執筆時点では、ページのタイトルが双子座モデルとなっています。）



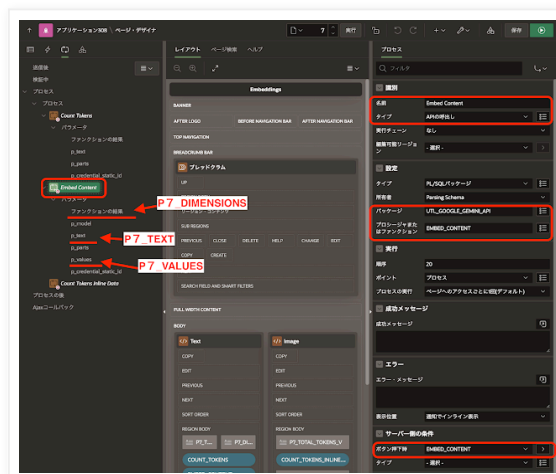
ボタンCOUNT_TOKENSのクリックで呼び出されるプロセスとして、Count Tokensを作成しています。パッケージUTL_GOOGLE_GEMINI_APIのファンクションCOUNT_TOKENSを呼び出しています。

ファンクションの結果にページ・アイテムP7_TOTAL_TOKENS、パラメータp_textにP7_TEXTを割り当てています。



ボタンEMBED_CONTENTのクリックで呼び出されるプロセスとして、Embed Contentを作成しています。パッケージUTL_GOOGLE_GEMINI_APIのファンクションEMBED_CONTENTを呼び出しています。

ファンクションの結果にページ・アイテムP7_DIMENSIONS、パラメータp_textにP7_TEXT、p_valuesにP7_VALUESを割り当てています。

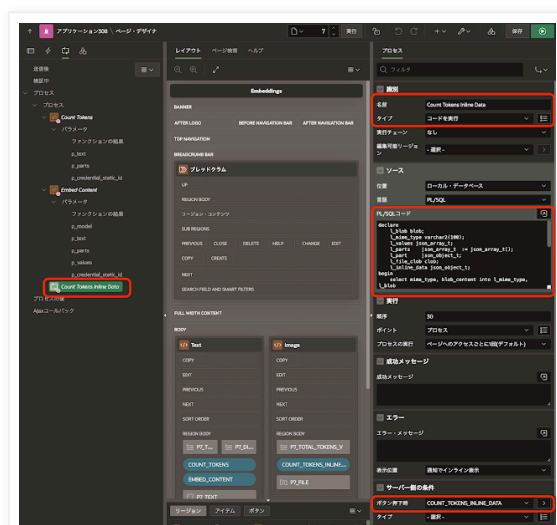


ボタンCOUNT_TOKENS_INLINE_DATAのクリックで呼び出されるプロセスとして、Count Tokens Inline Dataを作成しています。タイプとしてコードを実行を選択し、ソースのPL/SQLコードとして、以下を記述しています。

```
declare
    l_blob blob;
    l_mime_type varchar2(100);
    l_values json_array_t;
    l_parts    json_array_t := json_array_t();
    l_part     json_object_t;
    l_file_clob clob;
    l_inline_data json_object_t;
begin
    select mime_type, blob_content into l_mime_type, l_blob
    from apex_application_temp_files where name = :P7_FILE;
    l_part := json_object_t();
    l_inline_data := json_object_t();
    l_inline_data.put('mimeType', l_mime_type);
    l_file_clob := apex_web_service.blob2clobbase64(l_blob, 'N','N');
    l_inline_data.put('data', l_file_clob);
    l_part.put('inlineData', l_inline_data);
    l_parts.append(l_part);
    :P7_TOTAL_TOKENS_V := utl_google_gemini_api.count_tokens(
        p_parts => l_parts.to_clob()
        ,p_credential_static_id => :G_CREDENTIAL
    );
end;
```

count-tokens-inline-data.sql hosted with ❤ by GitHub

[view raw](#)



ページの作りは以上になります。

トークン数のカウントやベクトル埋め込みの生成を試してみます。

以下の文章のトークン数をカウントしてみます。夏目漱石の吾輩は猫であるの出だしです。

吾輩は猫である。名前はまだ無い。どこで生れたかとうと見当がつかぬ。何でも薄暗いじめじめした所でニャーニャー泣いていた事だけは記憶している。吾輩はここで始めて人間というものを見た。しかもあとで聞くとそれは書生という人間中で一番憚悪な種族であったそうだ。

文字数は126文字です。トークン数は70でした。

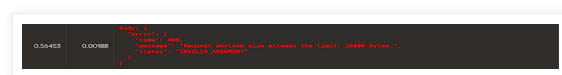
The screenshot shows the Hugging Face tokenizer interface. The 'Text' input field contains the Japanese text. The 'Total Tokens' field shows 70, and the 'Dimensions' field shows 768. The 'Count Tokens' button is highlighted. Below the text, the 'Embedding' field shows a list of 70 token embeddings, each a 768-dimensional vector.

Gemini Proの入力トークンの上限は30720なので、それを超えないように確認するために使用できます。

The screenshot shows the Hugging Face tokenizer interface. The 'Text' input field contains a longer Japanese text. The 'Total Tokens' field shows 31378, which is highlighted in red, indicating it exceeds the 30720 limit. The 'Dimensions' field shows 768. The 'Count Tokens' button is highlighted. Below the text, the 'Embedding' field shows a list of 31378 token embeddings, each a 768-dimensional vector.

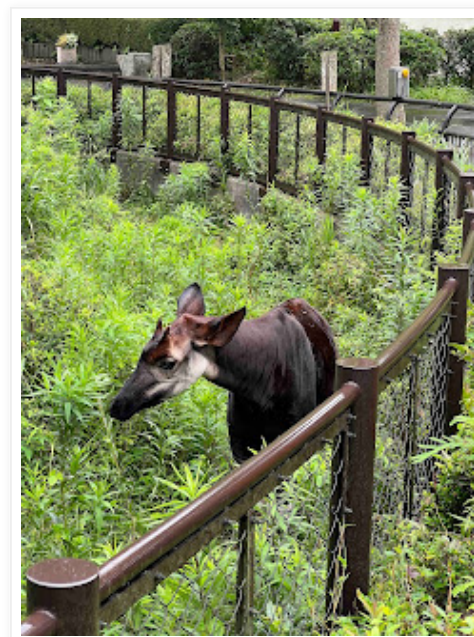
ベクトル埋め込みの生成については、よく資料を確認しておく必要がありました。長い文章からベクトル埋め込みを生成しようとすると、エラーが発生します。

エラー・メッセージを確認すると、embedContentの呼び出しでは、リクエストのペイロードのサイズが10000バイトに制限されているとのことでした。



モデルの説明を読むと「最大 2,048 トークンのテキスト用のエンベディングの作成用に最適化されています。」とのことでした。それぞれのベクトル埋め込みの生成では、日本語だと文章の長さを 3 千文字前後（大体10000バイトになるはず）に抑える必要がありそうです。

画像については、トークン数をカウントすることができました。以下の画像でトークン数は258でした。



Image

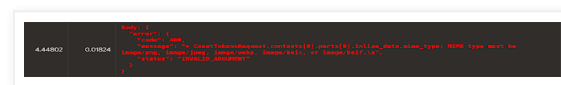
Total Tokens
258

Count Tokens

File

 ファイルの選択

動画のトークン数をカウントしてみたら、以下のエラーが発生しました。エラー・メッセージを読むと「MIME type must be image/png, image/jpeg, image/webp, image/heic, or image/heif.」なので、列記されているMIMEタイプに動画は含まれていません。



きちんと確認していませんが、`inlineData`や`fileData`は`mimetype`が画像または動画に限定されているため（[こちらのリファレンス](#)の`mimeType`に指定できる値が記載されています）、ベクトル埋め込みの生成にファイルを指定することはできませんでした。

これらの機能を追加したAPEXアプリケーションのエクスポートを以下に置きました。
<https://github.com/ujnak/apexapps/blob/master/exports/google-gemini3.zip>

Oracle APEXのアプリケーション作成の参考になれば幸いです。

完

[ウェブ バージョンを表示](#)

自己紹介

Yuji N.

日本オラクル株式会社に勤務していて、Oracle APEXのGroundbreaker Advocateを拝命しました。
こちらの記事につきましては、免責事項の参照をお願いいたします。

[詳細プロフィールを表示](#)

Powered by Blogger.
