

# 画像と文字が混在しているファイル

## ■この段落は文字情報あり

一見すると、すぐ下の段落と変わらないように見えますが、この段落は文字情報を持っています。ひらがなの「へ」とカタカナの「ヘ」でも、漢字の「夕」とカタカナの「タ」でも、文字として区別しながら検索可能です。検索はたいてい、Ctrl + Fのショートカットキーが使えます。覚えておくと便利です。検索した上で置換もしたいときは、Ctrl + Hのことが多いようです。

## ■この段落は画像

一見すると、上下の段落と変わらないように見えますが、この段落は画像です。文字情報は持っていません。ですので、ここからテキストを抽出しようとすると、一手間かかります。OCRは、どうしても精度100%とは行きません。字形の似た漢字もあれば、ひらがなの「へ」とカタカナの「ヘ」、漢字の「夕」とカタカナの「タ」のように、前後のコンテキストに頼らなければ弁別が難しい文字もありますよね。

## ■この段落は文字情報あり

この段落はまた、文字情報を持っています。ひらがなの「へ」とカタカナの「ヘ」でも、漢字の「夕」とカタカナの「タ」でも、文字として区別しながら検索可能です。