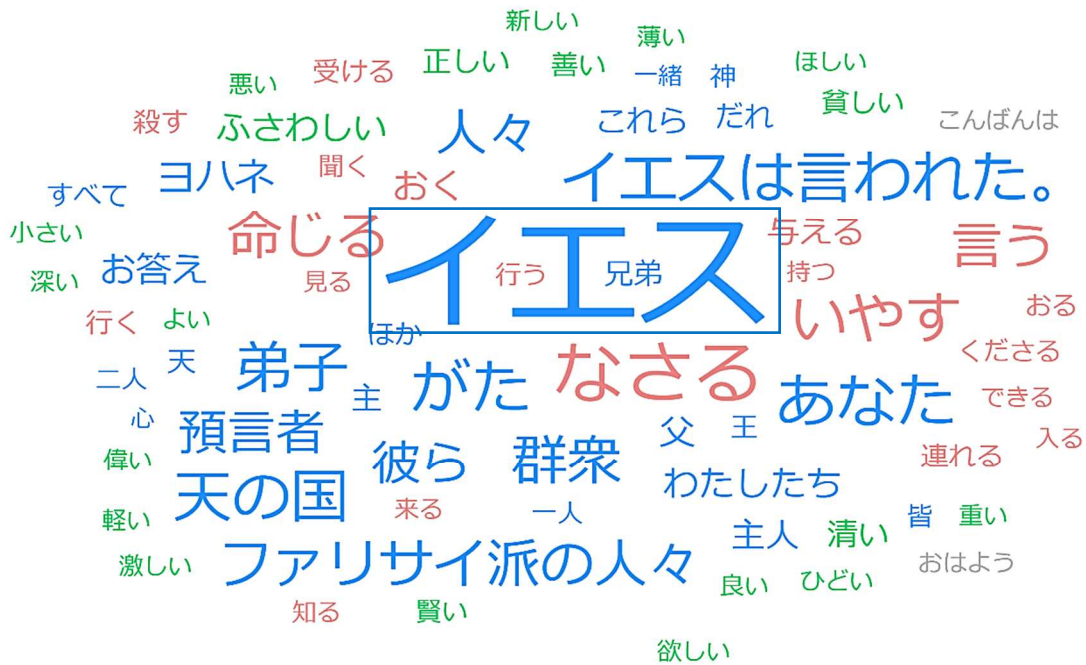
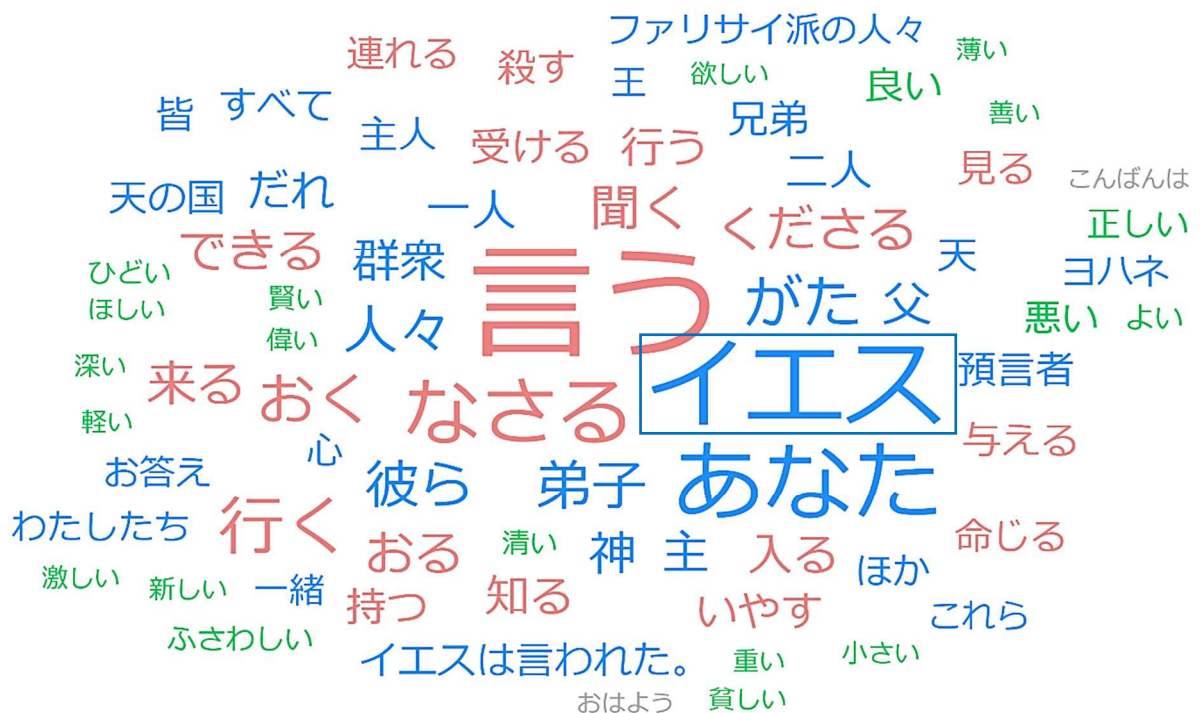


<1>マタイによる福音書(AIテキスト・マイニング結果)文字数:約 55,330 文字
イエス: 342 回登場する(イエスはマタイによる福音書に 342 回登場する)

▶スコア順: 単語の「重要度」を表す値順で、スコアが高い単語は、そのテキストを特徴づける単語です。



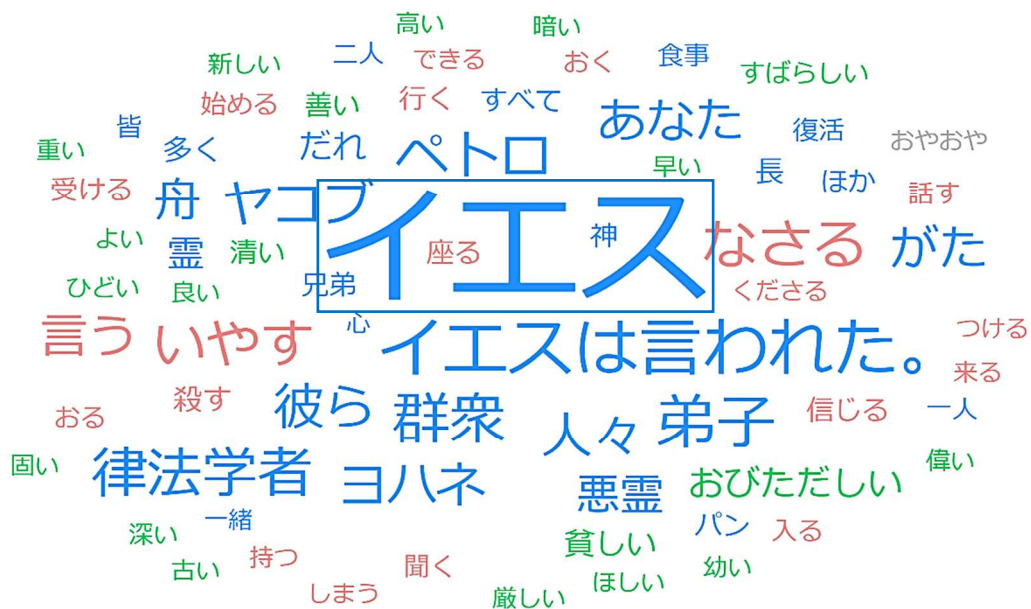
▶出現頻度順: 文章中に出現する単語の頻出度を表にしています。



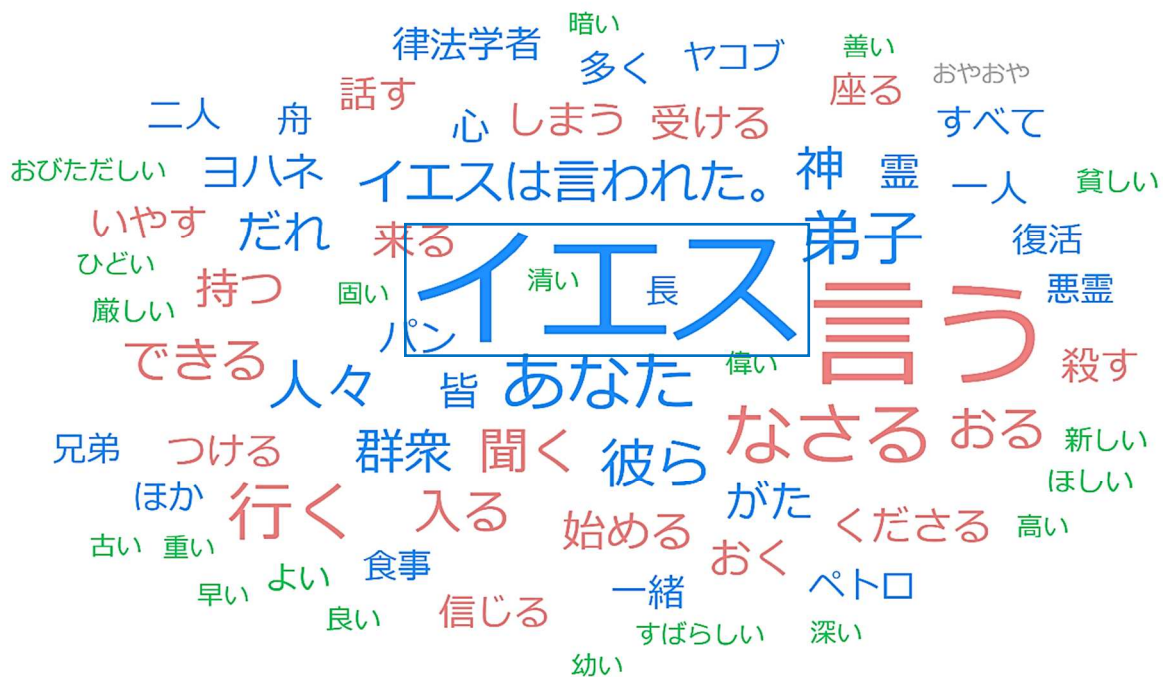
※上記の聖句は、[聖書研究ソフト「聖書 Navi Active」](#)により検索、抽出しています。
※ユーザーローカル [テキストマイニングツール](http://textmining.userlocal.jp/) (<http://textmining.userlocal.jp/>)による分析をしています(以下同じ)。

<2>マルコによる福音書(AIテキスト・マイニング結果)文字数:約 35,440 文字
イエス : 351 回登場する

▶スコア順

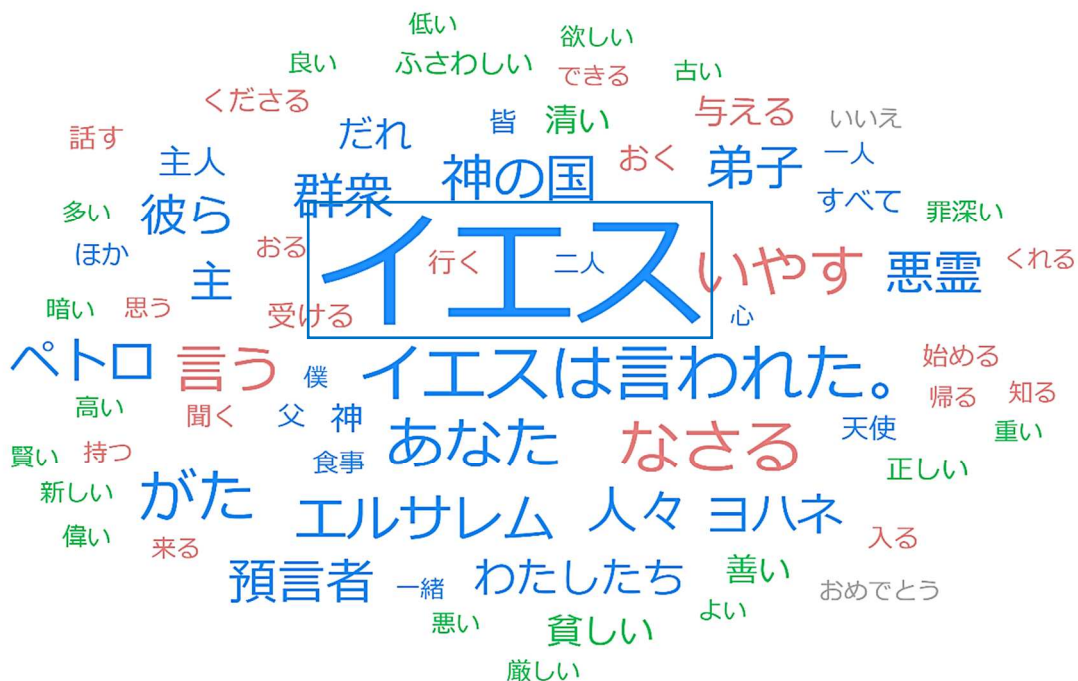


▶出現頻度順

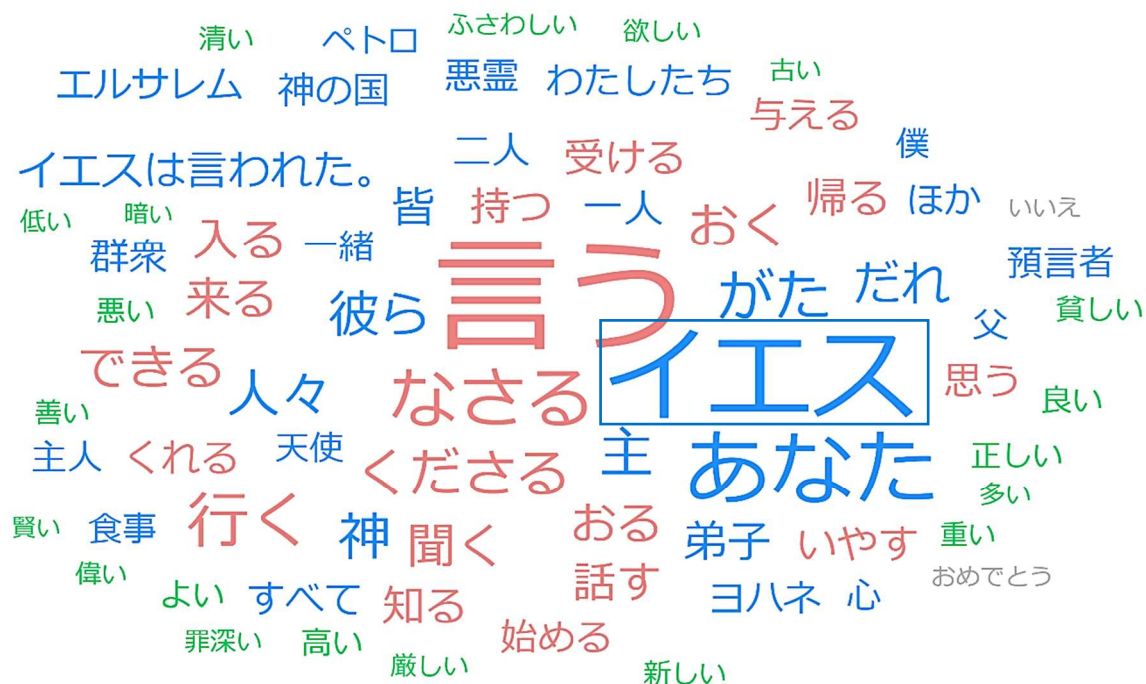


<3>ルカによる福音書(AI テキスト・マイニング結果) 文字数:約 59,900 文字
イエス: 412 回登場する

▶スコア順

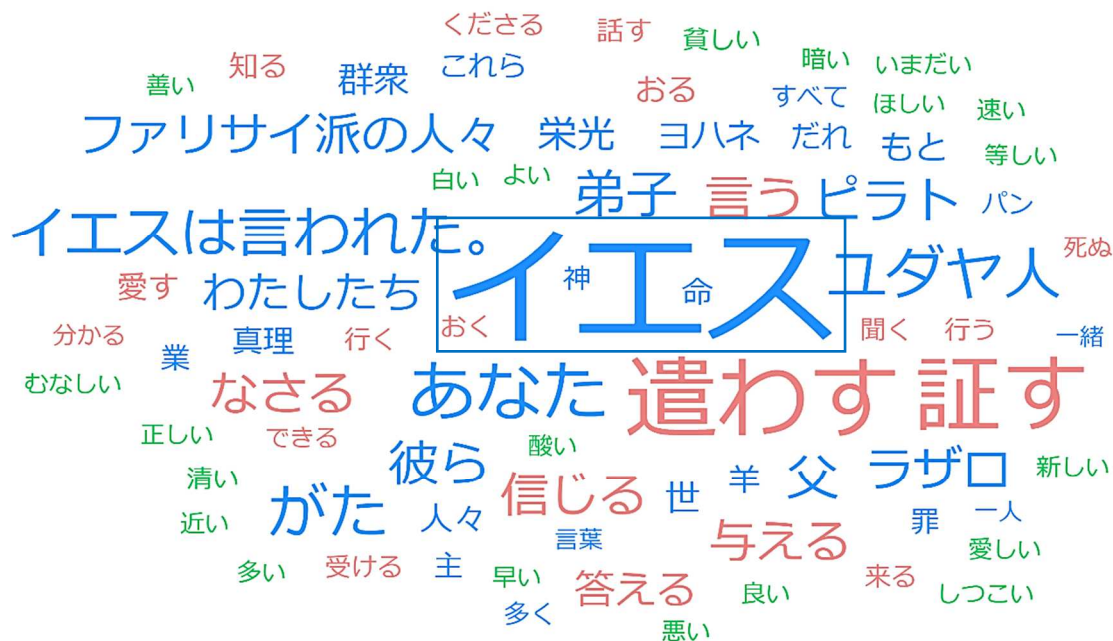


▶出現頻度順

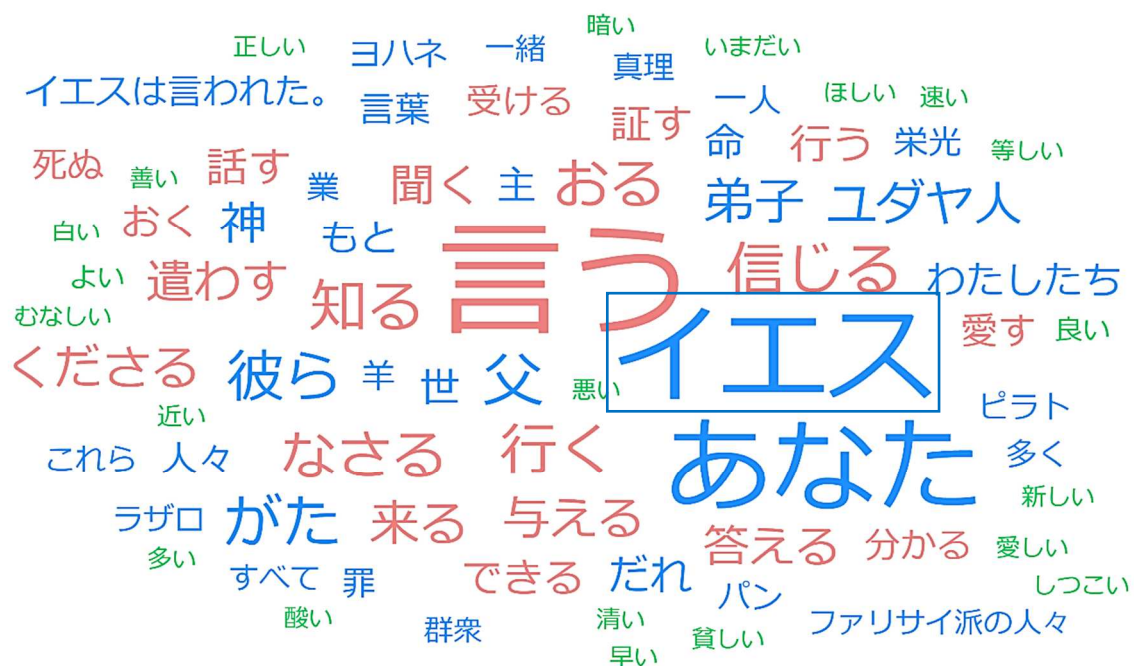


<4>ヨハネによる福音書(AIテキスト・マイニング結果)文字数:約 49,840 文字
イエス : 389 回登場する

▶スコア順



▶出現頻度順



▶ 「スコア」について

スコアは、その単語の「重要度」を表す値です。以下で、スコアがなぜ必要なのかと、その算出方法についてご紹介します。

一般的な文書では、単語の出現回数だけでいえば「今日」や「思う」「ある」などといった、“ごく一般的な単語”が何度も出現してしまいます。ただ、このような単語は、どういった文書にも出現する単語であるため、たとえ出現回数が多いとしても、意味が薄い、あまり重要ではない単語といえます。単純に回数だけをランキング化しても、一般的な語が混じってしまいその文章の特徴をつかむことができません。

この問題を解決するため、テキストマイニングでは、「一般的な文書でよく出る単語は、重要ではないため、重み付けを軽くする」、いっぽう「一般的な文書ではあまり出現しないけれど、調査対象の文書だけによく出現する単語は重視する」仕組みを取り入れています。

こういった特徴語を抽出するためのロジックとして、一般的に TF-IDF 法という統計処理をします。

→参考 URL <https://ja.wikipedia.org/wiki/Tf-idf>

この手法によって、出現回数だけでなく、重要度を加味した値が「スコア」です。スコアが高い単語は、そのテキストを特徴づける単語であるといえます。

出典・参考：※ユーザーローカル テキストマイニングツール (<http://textmining.userlocal.jp/>)