

## NEC Advanced Analytics Platform V1.5 マネージドサービス サイジングガイド (簡易版)

### テキスト分析編 (第1版)

NEC Advanced Analytics Platform V1.5 マネージドサービスでテキスト分析エンジンを使用して分析を行うにあたって各契約プランのメモリおよびストレージリソース量で扱うことが可能なデータ量(文数、文書数)の目安を表 1 に示します。

表 1 使用機能ごとの入力データ量と選択プランの目安

使用機能	入力データ量	選択するプラン
文クラスタリング	文数 13 万まで	エントリー エントリープラス(1 コア)
	文数 19 万まで	エントリープラス(4 コア)
	上記より多い	スタンダード
文書判別(学習処理)と 文書判別(判別処理)csv 出力	文書数 1.4 万まで	エントリー エントリープラス(1 コア)
	文書数 7.4 万まで	エントリープラス(4 コア)
	上記より多い	スタンダード
文書判別(学習処理)と 文書判別(判別処理)asf 出力	文書数 0.6 万まで	エントリー エントリープラス(1 コア)
	文書数 1.3 万まで	エントリープラス(4 コア)
	上記より多い	スタンダード

#### 前提条件：

- 文クラスタリングの入力データは 1 文あたり 128 バイト、文書判別の三つの機能の入力データは 1 文書あたり 1480 バイトを想定した数値となっています。
- ストレージについては、分析 10 回分の入力データと出力データを保存しておく前提での目安となっています。過去の入力データと出力データを 10 回分以上保存しておく場合は、各プランで扱える文数や文書数がより少なくなります。

#### 注意事項：

- 文クラスタリングに必要なメモリ量とストレージ量は文数の 2 乗オーダーで増加するので注意が必要です。
- asf (言語解析結果を含む JSON 形式) で出力する場合、オリジナルの入力文書の約 80 倍の容量が必要になるので注意が必要です。
- 分析で使用する入力データ以外のファイルがストレージにある場合、その分扱えるデータ量が減ります。
- データの独自の加工を行う場合、その加工において必要なファイル容量は上記目安値に含めていないため、加工ファイルがストレージを大きく使う場合は、扱えるデータ量が減ります。
- 本書で示す数値はあくまで目安ですので、お客様の分析データや分析方法によってはメモリ使用量やストレージ使用量が想定を上回る可能性があります。エントリープラン、エントリープラスプランでは後からメモリを追加できないため、およびエントリープランでは後からストレージを追加できないため、不安な場合は 1 ランク上のプランを選択される等ご検討ください。より確からしい見積もりを出すためには、本番運用と同等の特性を持つサンプルデータを用いるなどして実測することを推奨します。

#### 用語説明：

文数	分析対象の文の数。本書では 1 文あたり 128 バイトを想定しています。
文書数	分析対象の文書(テキストファイル)の数。本書では 1 文書あたり 1480 バイトを想定しています。
文クラスタリング	文書中の 1 文 (程度) に相当するテキストの集合を入力として、類似したテキストを同じグループ (クラスタと呼ぶ) にまとめ上げ、クラスタの集合を生成する機能です。
文書判別	機械学習の手法を使って、未知の文書を決められたカテゴリに分類する機能です。

#### 本書について：

- より詳細な見積もりが必要な場合は、「NEC Advanced Analytics Platform V1.5 マネージドサービス サイジングガイド」を参照してください。
- 本書を事業者の許可なくコピーおよびその配布、ホームページへの掲載を禁じます。