

## 日本語を解析する N-Gram For ALPHASHION の取扱説明書

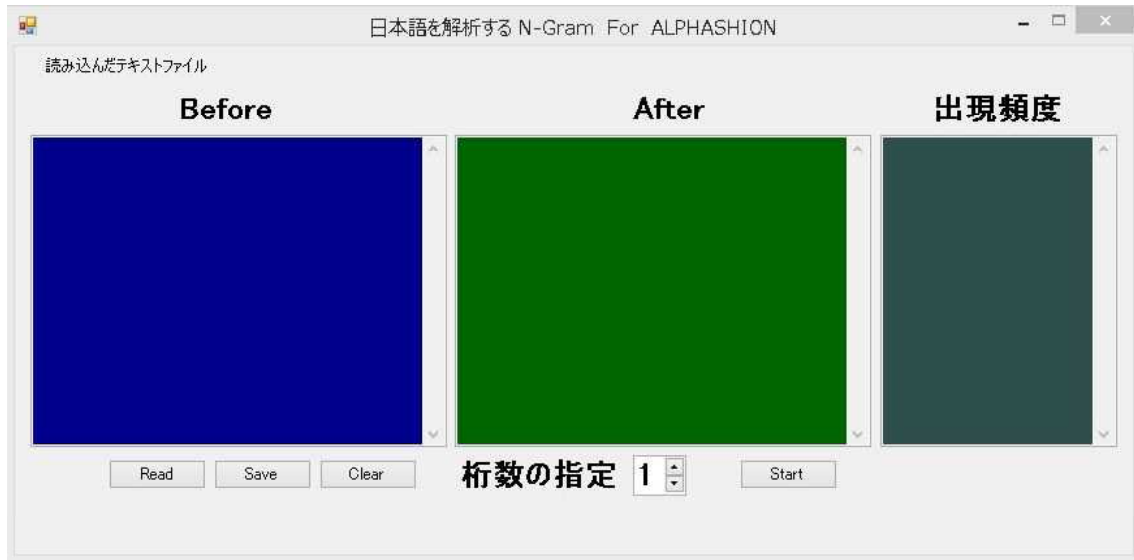
まずはダウンロードをして頂きましてありがとうございます。

本ソフトウェアを適切に取り扱って頂くための方法を説明していきます。

- 【 名 称 】 日本語を解析する N-Gram For ALPHASHION
- 【 目 的 】 文字の出現頻度のリストを作成。
- 【 開 発 日 】 2014年 2月 1日
- 【バージョン】 V e r 1.00
- 【 圧縮形式 】 z i p
- 【 著作権者 】 A L P H A S H I O N
- 【 変更履歴 】 2014年 2月 1日 V e r 1.00 完成
- 【 転載条件 】 中身を改変せず、非商用目的に限り転載が可能。
- 【問い合わせ】 万が一、気がついた点がありましたらメールでどうぞ。  
software@alphashion.info

- 【 補 足 】 夏目漱石先生の「坊ちゃん」を計算した結果付き。

## 取り扱いの手順。



起動をすると上記の画面が表示されます。

「Before」は、解析させたい文章を表示するエリアです。

「After」は、解析の途中を表示するエリアです。

「出現頻度」は、解析結果を表示するエリアです。

「桁数の指定」は、指定した桁数に基づいて分割し、降順に並び替えた上で表示されます。

指定できる数値は1～5の範囲です。

「Read」 → 対象のテキストファイルを読み込みます。(拡張子はTXTのみに対応)

「Save」 → 解析結果を保存します。

保存する際のファイルネームは変更することができます。

保存形式はタブ区切りのTXT形式です。表計算ソフトで読み込むことが可能になります。

「Clear」 → 読み込んだ文章を初期化します。

「Start」 → 解析を開始します。

解析したいテキストファイルは「B e f o r e」にドラッグ&ドロップ操作により読み込ませることもできます。

アンインストール手順。

本ソフトウェアは生成されたフォルダを削除することで消去が可能です。

レジストリファイルを操作しておりませんので、安心して削除が可能です。

#### 免責事項

プログラムは細心の注意を払って組んでいますが、計算の正確性は保障しておりません。

## 【具体的な操作手順について。】

まず、ブラウザのURL入力欄に青空文庫の夏目漱石の「坊ちゃん」のアドレスを打ち込むで表示をします。

<http://www.aozora.gr.jp/cards/000148/card752.html>

ここより、「XHTML ファイル | なし | 752\_14964.html」をクリックし、該当の小説を表示する。

文字列を選択しドラッグしながら範囲の選択を行う。

分らないから、私儀(わたくしぎ)都合有之(これあり)辞職の上東京へ歸り申候(もうしそる)につき左様御承知  
被下度候(さようごしょうちくたされたくそろ)以上とかいて校長宛(あて)にして郵便で出した。  
汽船(きせん)は夜六時の出帆(しゅっぱん)である。山嵐(さんらん)もおれも疲れて、ぐっぐう寝込んで眼が覚めたら、午後二時で  
あった。下女(かみめ)に巡査(じゅんさ)は来ないかと聞いたら参りませんと答えた。「赤シャツも野だも訴えなかったなあ」と二人は  
太きに笑った。  
その夜おれと山嵐(さんらん)はこの不浄(ふじょう)な地を離(はな)れた。船が岸を去れば去るほどいい心持ちがした。  
神戸から東京までは直行で新橋へ着いた時は、ようやく娑婆(しゃは)へ出たような気がした。山嵐(さんらん)とはすぐ分  
れたぎり今日まで逢う機会がない。  
清(きよ)の事を話すのを忘れていた。——おれが東京へ着いて下宿へも行かず、革靴(かはん)を提げたま  
ま、清(きよ)や帰ったよと飛び込んだら、あら坊(ぼく)ちゃん、よくまあ、早く帰って来て下さったと展(なみた)をほた(またと  
落した。おれもあまり嬉(うれ)しかったから、もう田舎(いなか)へは行かない、東京で清(きよ)とうちを持つんだと云っ  
た。  
その後ある人の周旋(しゅうせん)で街鉄(かいてつ)の技手(わざて)になった。月給は二十五円で、家賃は六円だ。清  
(は玄閻(げんかん)付きの家でなくても至極満足の様子であったが気の毒な事に今年の二月肺炎(はいえん)  
に罹(ひか)って死んでしまった。死ぬ前日おれを呼んで坊(ぼく)ちゃん後生(ごせい)だから清(きよ)が死んだら、坊(ぼく)ちゃんのお  
寺へ埋(う)めて下さい。お墓(はか)のなかで坊(ぼく)ちゃんの来るのを楽しみに待っておりますと云った。だから清(きよ)の墓(はか)は  
小日向(こひなた)の養源寺にある。

(明治三十九年四月)

マウスの右クリックより、「コピー」を選択しクリップボードに記憶させる。

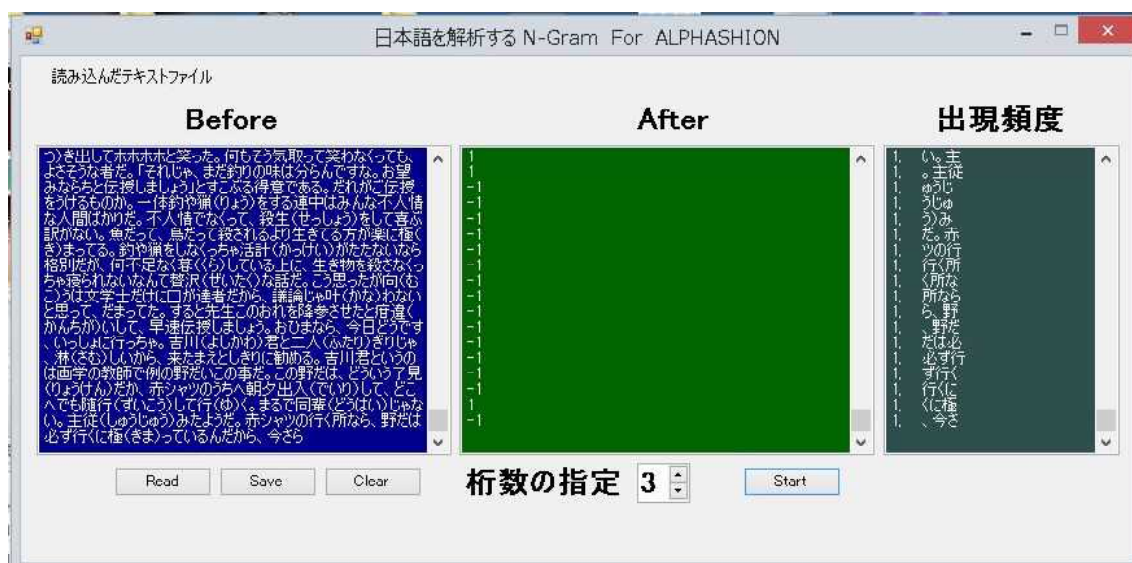
「Before」のエリア上でマウスの右クリックをし「貼り付け」を選択し小説を貼り付ける。

「桁数を指定する」は1～5の範囲で設定する。今回は3を指定する。

「Start」ボタンを押下する。

処理に時間がかかるため、しばらく待ちます。

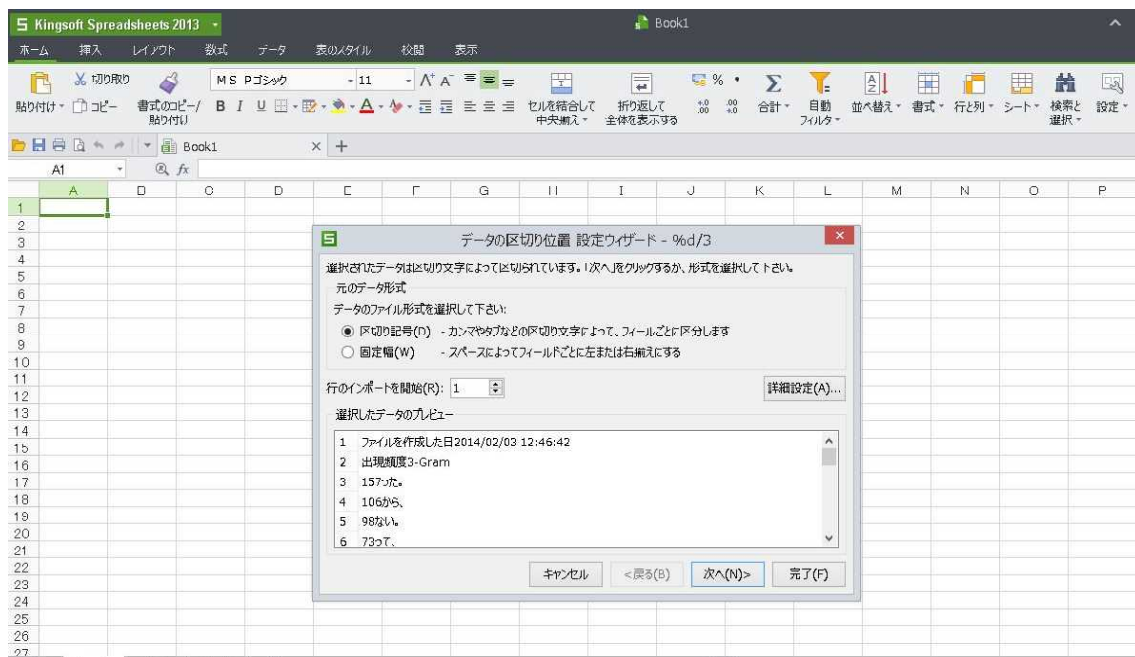
結果は以下のようになります。



解析結果を保存する場合は「Save」ボタンを押下する。

ファイル保存ダイアログが表示されたら任意のファイルネームにして保存する。  
ここでは「夏目漱石の坊ちゃん 3-Gram.txt」とする。

表計算ソフトで読み込みたい場合は、「タブ区切り」で読み込みます。



表計算ソフトにタブ区切りテキスト・ファイルを読み込ませたところ。

	A	B	C	D
1	ファイルを作成	2014/2/3 12:46		
2	出現頻度	3-Gram		
3	157	った。		
4	106	から、		
5	98	ない。		
6	73	って、		
7	60	ている		
8	60	ある。		
9	59	と云っ		
10	59	。おれ		
11	54	いた。		
12	54	した。		
13	54	おれは		
14	50	と思っ		
15	49	たら、		
16	46	ていた		
17	46	かった		
18	46	と云う		
19	44	云った		
20	41	いる。		
21	41	おれの		
22	39	らない		
23	39	たが、		
24	38	してい		
25	36	ばかり		
26	36	た。おれ		