KWIC Concordance for Windows

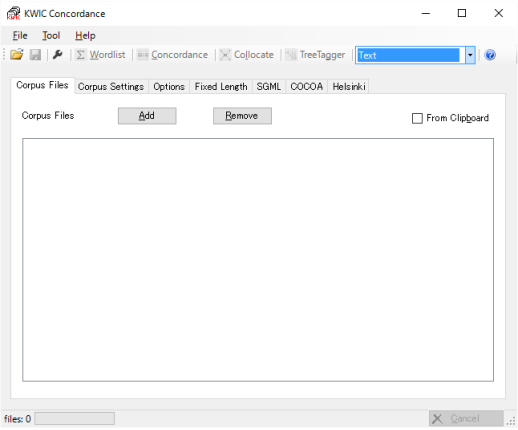
ヘルプ トピック

目次

- 使い方
- Menu
  - File
  - Tool
  - Help
- Corpus Type

使い方

コーパスファイルとして指定されたファイルを処理して、単語リスト・コンコーダンスに関するデータを新規ウインドウに処理して出力します。

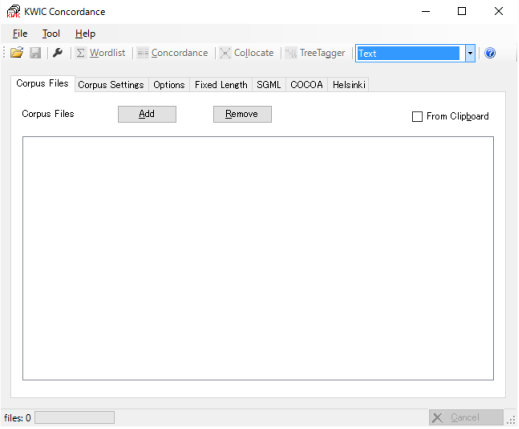


基本操作

はじめにデータファイルとしてのコーパスファイル を指定して下さい。処理は全てこのファイルに対して行われます。複数のファイルを指定できます。検索等の処理結果は、新規ウインドウに出力されます。行いたい処理をメニューから選び、処理して下さい。データの保存は、Copy により他のアプリケーションで保存、または Save 機能でテキストファイルに保存してください。

基本手順

- コーパスファイル選択（[Corpus Files] – [Add]/[Remove] ）

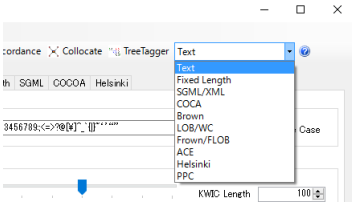


- Add Button：ファイル追加
- Remove Button：不要なファイルを除外
- ドラッグ＆ドロップでファイルを追加
  - From Clipboard:クリップボードにコピーしたテキストを一時的に扱い場合にチェックします。

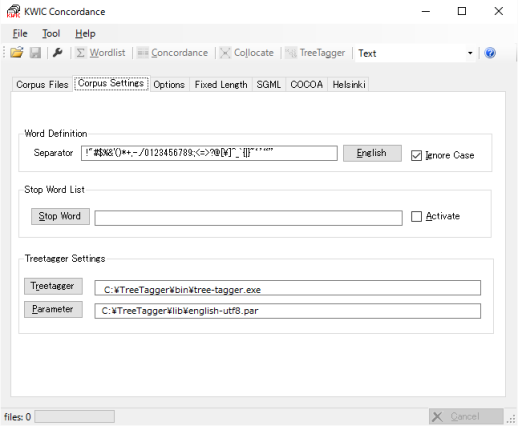
↓

- コーパスファイルに関する処理条件等の設定（[Corpus Type] / Options / 各オプション [Fixed Length] / [SGML] / [COCOA] / [Helsinki] ）

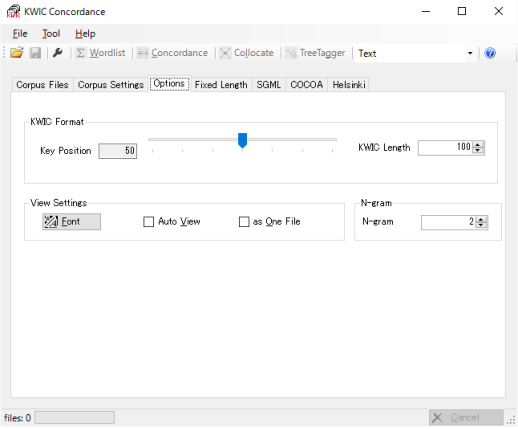
  - Corpus Type 選択



  - Corpus Settings 設定



  - Option 設定



- 各コーパス設定
  - Fixed Length
  - SGML
  - COCOA
  - Helsinki

↓

検索処理実行・表示(Tool)

- Wordlist
- Concordance
- Collocation
- N-gram
- TreeTagger

Menu

	File		Tool		Help
	Open	Σ	Wordlist		Contents
	Save		Concordance		About …
	Settings		Collocation		
	Option	N-g	N-gram		
	View		TreeTagger		
	Exit				

Toolbar

			Σ Wordlist	Concordance	Collocate	TreeTagger	Corpus Type	

Menu

File Menu

Open

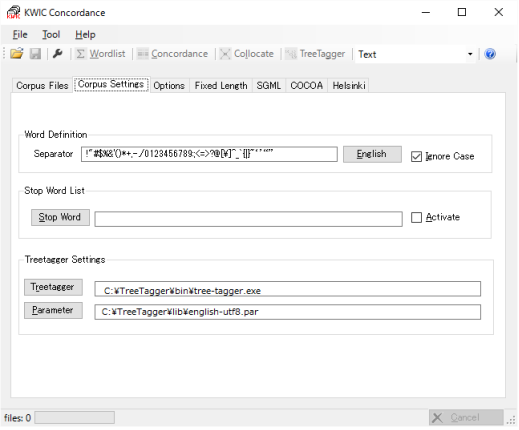
コーパスファイルとして登録された保存済みファイル一覧リスト読み込みます。
保存には[Save]コマンドを使用します。

Save

コーパスファイルとして登録されたファイル一覧リストを保存します。
再度コーパスファイルを復元するには[Open]コマンドを使用します。

Corpus Settings

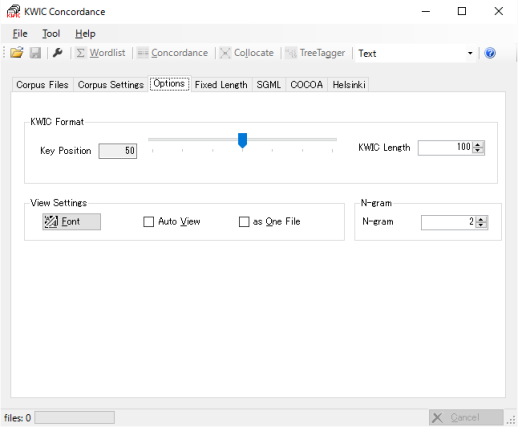
コーパス処理の際のコーパスの設定を行います。



Word Definition	
Separator	語の区切りとなる語以外の文字を指定します。[English]で英語の規定値を設定します。
Ignore Case	大文字・小文字を区別する指定をします。
Stop Word List	
Stop Word	Stop word の記録されたファイルを指定します。1 行 1 語で単語を指定します。 Activate をオンにすると、Wordlist と N-gram の場合に指定語を除外します。
Treetagger Setting	
Treetagger	TreeTagger 実行ファイル（tree-tagger.exe）を指定します。
Parameter	Parameter ファイル（English.par など）を指定します。

Option

コーパス処理の際の各種の設定を行います。

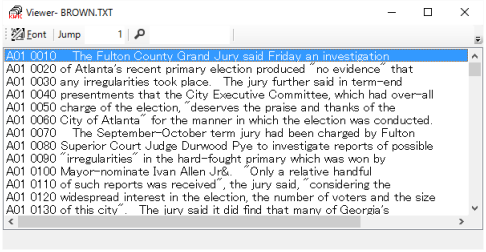


KWIC Format	
Key Position	KWIC Concordance 作成時のキーワードの位置を指定します。
KWIC Length	KWIC Concordance 作成時の 1 行の長さを指定します。
View Settings	
Font	表示フォントを指定します。
Auto View	KWIC Concordance 表示ウインドウからコーパスファイルへのクリックによる参照を可能にします。
As One File	複数ファイル指定時に、連続した一つのコーパスファイルとして扱う設定をします。
N-gram	
N-gram	N-gram の単位を指定します。(2～5)

View

簡易ファイル表示機能を起動します。

コーパスファイルなどの参照用に使用します。Font / Jump / Find 機能が使用できます。



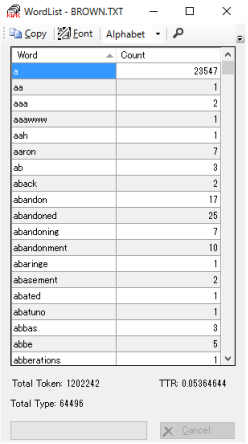
Exit

KWIC Concordance を終了します。

Tool Menu

Wordlist

[Tool] – [Wordlist] または Σ により、Word list を作成します。リストは見出し語、頻度、語尾順でソートできます。
Copy / Font / Sort / Find / Save 機能が使用できます。



Copy　単語リストをクリップボードにコピーします。

Font　単語リストの表示フォントを設定します。

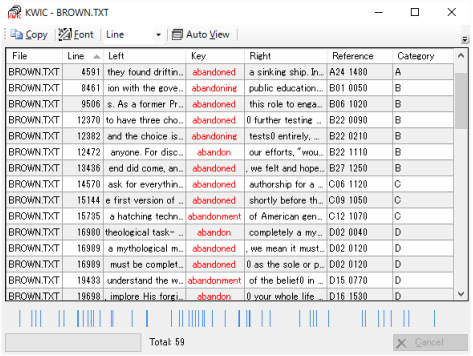
Sort　単語リストを単語（Alphabet）、頻度（Frequency）、語尾順（Tail）でソートします。

Find　文字列を検索します。

Save　単語リストをテキストファイルに保存します。

Concordance

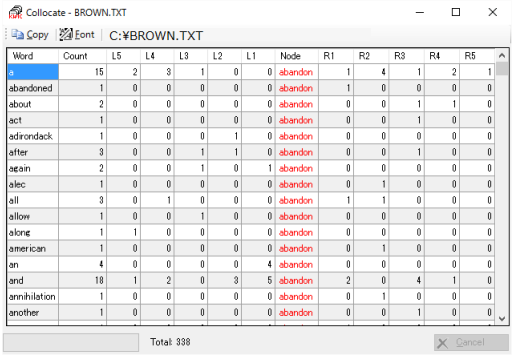
[Tool] – [Concordance] または により、Concordance list を作成します。リストは見出し語、行数、キーワード前後 5 語のいずれかでソートできます。
検索語のコーパス内での出現位置が下部に縦線でプロットされます。
Copy / Font / Sort / Auto View / Save 機能が使用できます。



Copy	コンコーダンスリストをクリップボードにコピーします。
Font	コンコーダンスリストの表示フォントを設定します。
Sort	コンコーダンスリストを行番号(Line)、キーワード(Key)、キーワード前後 5 語(L5-R5l)でソートします。
Auto View	簡易ファイル表示機能を起動します。
Save	コンコーダンスリストをテキストファイルに保存します。

Collocation

[Tool] – [Collocation] または により、Collocation list を作成します。リストは見出し語、頻度、コロケーション位置の頻度順でソートできます。
Copy / Font / Sort / Save 機能が使用できます。

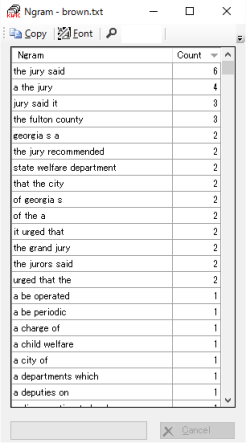


Word	Count	L5	L4	L3	L2	L1	Node	R1	R2	R3	R4	R5
a	15	2	3	1	0	0	abandon	1	4	1	2	1
abandoned	1	0	0	0	0	0	abandon	1	0	0	0	0
about	2	0	0	0	0	0	abandon	0	0	1	1	0
act	1	0	0	0	0	0	abandon	0	0	1	0	0
adirondack	1	0	0	0	1	0	abandon	0	0	0	0	0
after	3	0	0	1	1	0	abandon	0	0	1	0	0
again	2	0	0	1	0	1	abandon	0	0	0	0	0
alec	1	0	0	0	0	0	abandon	0	1	0	0	0
all	3	0	1	0	0	0	abandon	1	1	0	0	0
allow	1	0	0	1	0	0	abandon	0	0	0	0	0
along	1	1	0	0	0	0	abandon	0	0	0	0	0
american	1	0	0	0	0	0	abandon	0	1	0	0	0
an	4	0	0	0	0	4	abandon	0	0	0	0	0
and	18	1	2	0	3	5	abandon	2	0	4	1	0
annihilation	1	0	0	0	0	0	abandon	0	1	0	0	0
another	1	0	0	0	0	0	abandon	0	0	1	0	0

Copy	コロケーションリストをクリップボードにコピーします。
Font	コロケーションリストの表示フォントを設定します。
Sort	コロケーションリストを単語(Word)、頻度(Count)、キーワード(Node)L5 – R5 のヘッダーをクリックすることでソートします。
Save	コロケーションリストをテキストファイルに保存します。

N-gram

[Tool] – [N-gram] により、N-gram List を作成します。N の値は[Options]にて設定してください。
Copy / Font / Find / Save 機能が使用できます。



Copy　　N-gram リストをクリップボードにコピーします。

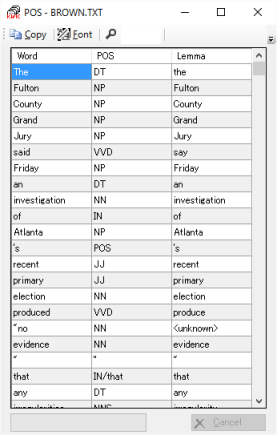
Font　　N-gram リストの表示フォントを設定します。

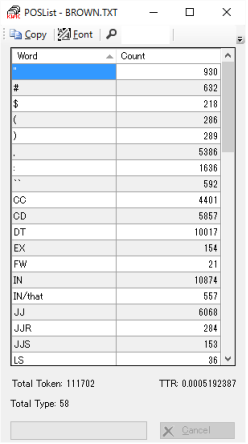
Find　　文字列を検索します。

Save　　N-gram リストをテキストファイルに保存します。

Treetagger

[Tool] – [Treetagger] または により、POS（品詞、レンマリスト）を作成します。作成された POS リストから、レンマリスト、POS 集計リストも併せて作成します。
TreeTagger の機能を内部より起動します。TreeTagger は事前にインストールする必要があります。Treetagger のサイト（http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/）より Windows 版 TreeTagger および Parameter ファイルをダウンロードしインストールしてください。Perl をインストールする必要はありません。インストールしたディレクトリを[Options]にて設定してください。
コマンドプロンプトをチャイルドプロセスとして起動しますので、処理に多少の時間を要します。
適切なインストールがなされていないと、正常に機能しない可能性があります。
Copy / Font / Find / Save 機能が使用できます。





Copy	POS リストをクリップボードにコピーします。
Font	POS リストの表示フォントを設定します。
Find	文字列を検索します。
Save	POS リストをテキストファイルに保存します。

Options

各種オプションを設定します。

Corpus Settings

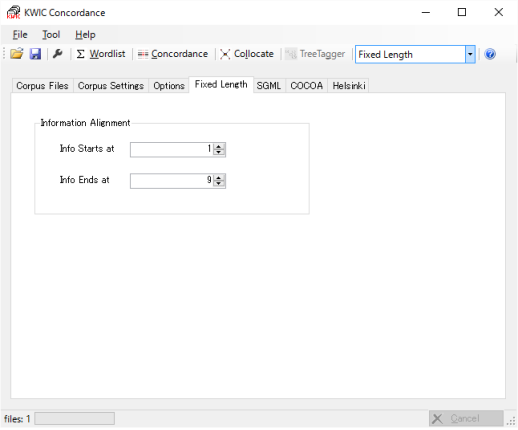
コーパスの設定を行います。

Options

KWIC Concordance の設定を行います。

Fixed Length

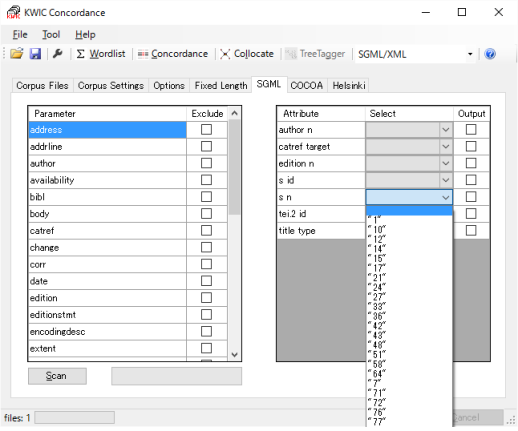
固定長形式のコーパスに関する設定を行います。



Info Starts at	参照部開始桁位置　（1～99）
Info Ends at	参照部終了桁位置　（2～100）

SGML

SGML/XML 形式のコーパスに関する設定を行います。

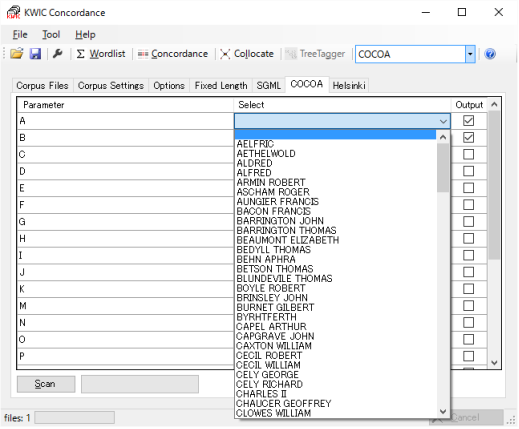


Parameter	
Parameter	<　　>で示されるタグを表示します。

Exclude	該当するタグを選択した状態で、Exclude チェックボックスを On にするとその部分を検索対象から除外します。
Attribute	
Attribute	<>で示されるタグの内、＝記号で示される属性(attribute)を含むタグを表示します。
Select	カテゴリー属性の条件をリストから選択して下さい。この条件にあったもののみが処理対象になります。対象を制限しない場合は、ブランク(空白)を選んでください。
Output	該当するタグを選択した状態で、Output チェックボックスを On にするとその部分の属性の情報を用例と共に出力します。
Scan	
Scan	コーパスファイルに登録されているコーパスファイルを調べて Parameter, Attribute それぞれの項目を設定します。

COCOA

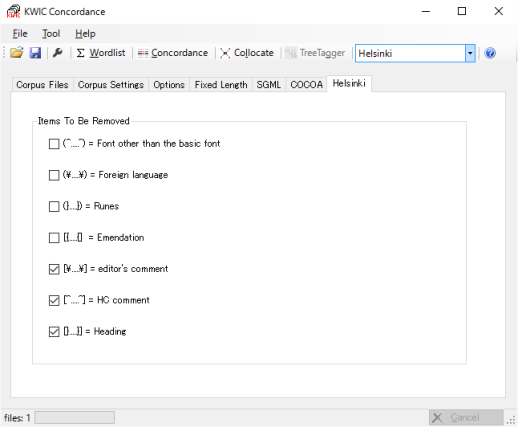
COCOA 形式のコーパスに関する設定を行います。



Parameter	リストボックスから指定する COCOA 形式のカテゴリーのコード。
Select	カテゴリーのコードの条件をリストから選択して下さい。この条件にあったもののみが処理対象になります。対象を制限しない場合は、ブランク(空白)を選んでください。(例下記では例えば、A カテゴリーに AUNGIER FRANCIS などの条件が付されたテキスト部分のみが、単語リスト、コンコーダンスなどの対象になります。)
Output	カテゴリーの情報を出力するかどうかを指定します。チェックボックスを On にするとそのカテゴリの情報を用例と共に出力します。(KWIC のみ有効)
Scan	Corpus Files に設定されているファイルから、Parameter とその属性値を読み取り、リストに設定します。

Helsinki

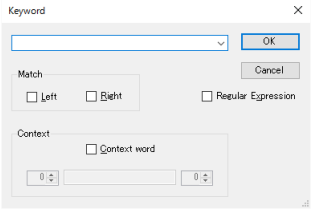
Helsinki コーパスに関する設定を行います。COCOA 形式の設定も併せて行います。
Helsinki Corpus で指定されている以下の Text level のコードを、チェックボックスを On にすることによって処理対象から除外できます。



(^ ^)	= font other than the basic font
(¥ ¥)	= foreign language
[]]]	= runes
[[[]	= emendation
[¥ ¥]	= editor's comment（ディフォルト On）
[^ ^]	= Helsinki Corpus comment（ディフォルト On）
[]]]	= heading（ディフォルト On）

Keyword Dialog

コンコーダンス、およびコロケーション作成時に表示され、検索語を指定します。



Keyword	コーパス処理のための文字列を入力して下さい。大文字・小文字の区別はオプションの指定に従います。過去の履歴を参照可能。
Match	前方一致・後方一致した文字列を処理対象にするかの指定します。
Regular Expression	検索する文字列に正規表現を使用する場合には、このチェックボックスを On にします。正規表現の詳しい設定については、正規表現をご覧下さい。
Context	コンコーダンスあるいはコロケーションリスト作成時に、このチェックを On にする事により、処理対象となる行に制限を加えます。コンテキストのチェックを ON の後、下に位置するエディットボックスに共起する文字列を指定して下さい。さらに、その指定文字列が、キーワードの左右何語目までの範囲に共起するかの指定をして下さい。指定文字列を指定範囲内に含むもののみを処理対象として、コンコーダンスあるいはコロケーションを作成します。Regular Expression（正規表現）指定時には、正規表現による文字列の指定も行えます。

正規表現

以下のような正規表現の設定ができます。

文字クラス	照合する部分文字列を定義する正規表現文字のセット
量指定子	正規表現の動作を変更する、量データの指定
グループ	正規表現で部分式のグループ
文字のエスケープ	文字が演算子ではなく、照合対象文字として解釈する必要があることを通知するエスケープ文字のセット

文字クラス

[character_group]	(文字範囲) 文字範囲内にある任意の文字に一致します。 文字範囲は連続する一連の文字で、範囲の最初の文字、ハイフン (-)、および範囲の最後の文字を指定することで定義されます。2 つの文字の Unicode コード ポイントが隣接している場合、それらの文字は連続しています。複数の文字範囲を連結することもできます。 たとえば、"0" ～ "9" の範囲の 10 進数、"a" ～ "f" の範囲の小文字、および "A" ～ "F" の範囲の大文字を指定するには、[0-9a-fA-F] を使用します。
[^ character_group]	(文字範囲) 文字範囲内にある任意の文字に一致します。 文字範囲は連続する一連の文字で、範囲の最初の文字、ハイフン (-)、および範囲の最後の文字を指定することで定義されます。2 つの文字の Unicode コード ポイントが隣接している場合、それらの文字は連続しています。複数の文字範囲を連結することもできます。 たとえば、"0" ～ "9" の範囲の 10 進数、"a" ～ "f" の範囲の小文字、および "A" ～ "F" の範囲の大文字を指定するには、[0-9a-fA-F] を使用します。
[firstCharacter-lastCharacter]	(文字範囲) 文字範囲内にある任意の文字に一致します。 文字範囲は連続する一連の文字で、範囲の最初の文字、ハイフン (-)、および範囲の最後の文字を指定することで定義されます。2 つの文字の Unicode コード ポイントが隣接している場合、それらの文字は連続しています。複数の文字範囲を連結することもできます。 たとえば、"0" ～ "9" の範囲の 10 進数、"a" ～ "f" の範囲の小文字、および "A" ～ "F" の範囲の大文字を指定するには、[0-9a-fA-F] を使用します。
.	(文字範囲) 文字範囲内にある任意の文字に一致します。 文字範囲は連続する一連の文字で、範囲の最初の文字、ハイフン (-)、および範囲の最後の文字を指定することで定義されます。2 つの文字の Unicode コード ポイントが隣接している場合、それらの文字は連続しています。複数の文字範囲を連結することもできます。 たとえば、"0" ～ "9" の範囲の 10 進数、"a" ～ "f" の範囲の小文字、および "A" ～ "F" の範囲の大文字を指定するには、[0-9a-fA-F] を使用します。
¥w	単語に使用される任意の文字と一致します。Unicode 一般カテゴリ [¥p{Ll}¥p{Lu}¥p{Lt}¥p{Lo}¥p{Nd}¥p{Pc}¥p{Lm}] と同じになります。ECMAScript オプションによって ECMAScript 準拠の動作が指定された場合、¥w は [a-zA-Z_0-9] と同じになります。
¥W	単語に使用される文字以外の任意の文字と一致します。Unicode 一般カテゴリ [^¥p{Ll}¥p{Lu}¥p{Lt}¥p{Lo}¥p{Nd}¥p{Pc}¥p{Lm}] と同じになります。ECMAScript オプションによって ECMAScript 準拠の動作が指定された場合、¥W は [^a-zA-Z_0-9] と同じになります。

¥s	空白文字と一致します。エスケープ シーケンスおよび Unicode 一般カテゴリ [¥f¥n¥r¥t¥v¥x85¥p{Z}] と同じになります。ECMAScript オプションによって ECMAScript 準拠の動作が指定された場合、¥s は [¥f¥n¥r¥t¥v] と同じになります。
¥d	10 進数字と一致します。Unicode の ¥p{Nd} および ECMAScript の動作である非 Unicode の [0-9] と同じになります。
¥D	数字以外と一致します。Unicode の ¥P{Nd} および ECMAScript の動作である非 Unicode の [^0-9] と同じになります。

量指定子

*	ゼロ回以上の一致を指定します。¥w*、(abc)* など。{0,} と等価です。
+	1 回以上の一致を指定します。¥w+、(abc)+ など。{1,} と等価です。
?	ゼロ回または 1 回の一致を指定します。¥w?、(abc)? など。{0,1} と等価です。
{n}	ちょうど n 回の一致を指定します。(pizza){2}
{n,}	n 回以上の一致を指定します。(abc){2,}
{n,m}	n 回以上 m 回以下の一致を指定します。
*?	できるだけ少ない繰り返しを処理する最初の一致を指定します (最短一致の * と同じ)。
+?	1 回以上でできるだけ少ない繰り返しを指定します (最短一致の + と同じ)。
??	できればゼロ回、できない場合は 1 回の繰り返しを指定します (最短一致の ?)。
{n}?	{n} (lazy {n}) と等価です。
{n,}?	n 回以上でできるだけ少ない繰り返しを指定します (最短一致の {n,})。
{n,m}?	n 回以上 m 回以下でできるだけ少ない繰り返しを指定します (最短一致の {n,m})。

グループ

|	(縦棒) 文字で区切られた用語 (たとえば cat|dog|tiger) のいずれかに一致します。最も左で一致したものが優先されます。
( )	一致した部分式

文字のエスケープ

通常の文字	.$ ^ { [(|) * + ? ¥ 以外の文字は、それ自体に一致します。
¥	エスケープ文字として認識されない文字が後ろに付いている場合は、その文字と一致します。たとえば、¥* は ¥x2A と同じです。
メモ	エスケープ文字 ¥b は、特殊なケースです。正規表現の場合、¥b はワード境界 (¥w 文字と ¥W 文字の間) を表しますが、[] 文字クラスの内側では、¥b はバックスペース文字を表します。置換パターンの場合、¥b は常にバックスペースを表します。

例

B*	空列、B、BB など 0 個以上の文字列に適合
AB*C	AC、ABC、ABBC などに適合

AB+C	ABC、ABC、ABBBC などに適合
ABB*C	ABC、ABBC、ABBC などに適合
AB?C	AC または ABC に適合
A|B	A または B のどちらかの文字列に適合
[AB]	A または B の文字に適合
[A-Z]	任意の英大文字 1 文字の文字列に適合
[A-Za-z]	任意の英大・小文字 1 文字の文字列に適合
[A-Z]*	任意の英大文字からなる 0 文字以上の文字列に適合
[A-Z]+	任意の英大文字からなる 1 文字以上の文字列に適合
[AB]?	A または B の 0 または 1 文字の文字列に適合
[^AB]	A または B でない文字列に適合
[^A-Z]	A から Z でない文字列に適合
(AB)+C	ABC、ABABC、ABABABC などに適合

詳細については Microsoft サイト 「正規表現言語要素」をご覧ください。
または、正規表現の詳しい説明は、A. V. エイホ、 B.W.カーニハン、P.J.ワインバーガー「プログラミング言語 AWK」(トッパン)、「詳説　正規表現」(オーム社)等をご参照ください。

Help Menu

Contents

ヘルプファイル（このファイル）を表示します。

About

バージョン、著作権を表示します。
KWIC Concordance for Windows Ver. 5.4 © Satoru Tsukamoto (Nihon University)

Corpus Type

Text	参照部を持たないテキストファイル
Fixed Length	参照部の桁位置の固定された固定長形式のファイル
SGML	<A> ... </A> <B> ... </B>の形式で参照部が設定されている SGML 形式のファイル
COCOA	<A > <B >の形式で参照部が設定されている COCOA 形式のファイル
Brown	The Brown Corpus (The Standard Corpus of Present-Day Edited American English)(Fixed Length)
LOB / WC	The LOB Corpus (The Lancaster-Oslo/Bergen Corpus) The Wellington Corpus of Written New Zealand English (WC) (Fixed Length)
Frown / FLOB	The Freiburg-LOB Corpus of British English (FLOB) The Freiburg-Brown Corpus of American English (FROWN) (Fixed Length)
ACE	The Australian Corpus of English (SGML)
Helsinki	Helsinki Corpus 専用(COCOA 形式＋独自コメント)
PPC	Penn-Helsinki Parsed Corpus psd ファイル専用

Text の例

一般的な参照部を持たないテキスト。

Pride and Prejudice
by Jane Austen

Chapter 1

It is a truth universally acknowledged, that a single man in possession of a large fortune must be in want of a wife.

However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered the rightful property of someone or other of their daughters.

Fixed Length の例

```
101010000010 WHAN THAT APRILL WITH HIS SHOURES SOOTE
201010000020 THE DROGHTE OF MARCH HATH PERCED TO THE ROOTE,
301010000030 AND BATHED EVERY VEYNE IN SWICH LICOUR
401010000040 OF WHICH VERTU ENGENDRED IS THE FLOUR;
501010000050 WHAN ZEPHIRUS EEK WITH HIS SWEETE BREETH
601010000060 INSPIRED HATH IN EVERY HOLT AND HEETH
701010000070 TENDRE CROPPES, AND THE YONGE SONNE
801010000080 HATH IN THE RAM HIS HALVE COURS YRONNE,
901010000090 AND SMALE FOWELES MAKEN MELODYE,
```

1001010000100 THAT SLEPEN AL THE NYGHT WITH OPEN YE
1101010000110 (SO PRIKETH HEM NATURE IN HIR CORAGES);
1201010000120 THANNE LONGEN FOLK TO GOON ON PILGRIMAGES,
1301010000130 AND PALMERES FOR TO SEKEN STRAUNGE STRONDES,
1401010000140 TO FERNE HALWES, KOWTHE IN SONDRY LONDES;

SGML の例

<TEI.1 ID=T22010 n=B17.1>
<TEI.header><file.description><title.statement><title>ChronA</title></title.statement><extent.statement> 637 cits</extent.statement><publication.statement> TEI OE Corpus Release 1</publication.statement><source.description><citn> The Parker Chronicle: Plummer 1892-9. </citn></source.description></file.description>
<encoding.declarations><reference.system> Cited by year following ed.; long entries are also cited by line no. assigned by DOE, following lineation of ed.
</reference.system><text.category def=PIS></encoding.declarations> </TEI.header>
<TEXT><s id="T22010000100" n="0.1">&t;Y GEARE &t;E W&a;S AGAN FRAM CRISTES ACENnesse <foreign> cccc</foreign> wintra. & <foreign>xciii</foreign> uuintra. &t;a Cerdic & Cynric his sunu cuom up &a;t Cerdices oran. mid <foreign>v</foreign> scipum. & se Cerdic w&a;s. Elesing. Elesa. Esling. Esla. Gewising. Giwis. Wiging. Wig. Freawining. Freawine. Fri&t;ugaring. Fri&t;ugar. Bronding. Brond B&a;ld&a;ging. B&a;ld&a;g Wodening.</s>
<s id="T22010000200" n="0.7">Ond &t;&a;s ymb <foreign>vi</foreign> gear &t;&a;s &t;e hie upcuomon geeodon West Seaxna rice, & &t;&a;t uu&a;run &t;a &a;restan cyningas &t;e West Seaxna lond on Wealum geeodon; & he h&a;fde &t;&a;t rice <foreign> xvi</foreign> gear, & &t;a he gefor &t;a feng his sunu Cynric to &t;am rice, & heold <foreign>xvii</foreign> winter.</s>
<s id="T22010000300" n="0.11">&t;a he gefor &t;a feng Ceol to &t;am rice, & heold <foreign>vi</foreign> gear.</s> </TEXT></tei.1>

HTML 形式のファイルは、この SGML の一種のため、SGML 形式に設定することにより処理できます。ただし、アトリビュート(属性)の情報・値が正確に付加されない場合があります。

The British National Corpus (BNC)の例

<bncDoc id=BDAPJ n=PetoIn> <header creator=dominic status=UPDATE update="2 Feb 1999" type=TEXT> <fileDesc> <titStmt> <title>Report on visit to Peto Institute -- an electronic version</title> <respStmt> <resp> Data capture and transcription </resp> <name> Oxford University Press </name> </respStmt> </titStmt> … <s n="0001"><w NP1>Peto <w NN1>Institute<c YCOM>, <w NP1>Budapest<c YCOM>, <w NP1> Hungary<c YSTP>. </s><s n="0002"><w MCMC>3-14 <w NPM1>December <w MC>1990 <w II> by <w NP1>J <w NP1>Davis <w CC>and <w NP1>J <w NP1>Eccleshall<c YCOM>, <w NP1> Mabel <w NP1>Prichard <w NNL1>School<c YCOM>, <w NP1>Oxford<c YSTP>. </s> <div1 complete=Y org=SEQ> <s n="0003"><w NN1>Introduction </s>
<s n="0004"><w II>In <w NPM1>November <w MC>1990<c YCOM>, <w AT>the <w MD>First <w NN1> World <w NN1>Congress <w IO>of <w AT>the <w JJ>International <w NP1>Petö <w NN1> Institute <w VBDZ>was <w VVN>held <w II>in <w NP1>Budapest<c YCOM>, <w NP1> Hungary<c YSTP>. </s><s n="0005"><w DA2>Several <w NNO>hundred <w NN>people <w VVD> attended <w II>from <w MC>37 <w JJ>different <w NN2>countries<c YSTP>. </s> <s n="0006"><w JJ>Various <w NN2>disciplines <w VBDR>were <w VVN>represented <w II> including <w NN2>educationalists<c YCOM>, <w

NN2>psychologists<c YCOM>, <w NN2> doctors <w CC>and <w NN2>others <w II>in <w JJ>associated <w JJ>medical <w NN2> professions<c YSTP>. </s>
<s n="0007"><w AT>The <w NN1>Congress <w VVD>lasted <w MC>three <w NNT2>days<c YCOM>, <w CC> and <w VBDZ>was <w VVN>followed <w II>by <w AT1>a <w MC>two <w NNT1>week <w NN1> Information <w NN1>Course <w II>at <w AT>the <w NN1>Institute<c YSTP>. </s> <s n="0008"><w PPIS2>We <w VBDR>were <w MC>two <w IO>of <w AT>the <w MC>seventy <w NN> people <w PNQS>who <w VVD>took <w NN1>part <w II>in <w DD1>this <w NN1>course<c YSTP>. </s>

COCOA の例

<P shakespeare>
<T Twelfth Night>
<S 1>
<A 1>
((A city in Illyria, and the sea-coast near it.))
((DUKE ORSINO's palace.))
((#Enter DUKE ORSINO, CURIO, and other Lords; Musicians attending))
<C DUKE ORSINO>
If music be the food of love, play on; Give me excess of it, that, surfeiting, The appetite may sicken, and so die. That strain again! it had a dying fall: O, it came o'er my ear like the sweet sound, That breathes upon a bank of violets, Stealing and giving odour! Enough; no more: 'Tis not so sweet now as it was before. O spirit of love! how quick and fresh art thou, That, notwithstanding thy capacity Receiveth as the sea, nought enters there, Of what validity and pitch soe'er, But falls into abatement and low price, Even in a minute: so full of shapes is fancy That it alone is high fantastical.
<C CURIO> Will you go hunt, my lord?
<C DUKE ORSINO> What, Curio?
<C CURIO> The hart.
<C DUKE ORSINO> Why, so I do, the noblest that I have: O, when mine eyes did see Olivia first, Methought she purged the air of pestilence! That instant was I turn'd into a hart; And my desires, like fell and cruel hounds, E'er since pursue me.

Brown の例

A01 0010 The Fulton County Grand Jury said Friday an investigation
A01 0020 of Atlanta's recent primary election produced "no evidence" that
A01 0030 any irregularities took place. The jury further said in term-end
A01 0040 presentments that the City Executive Committee, which had over-all
A01 0050 charge of the election, "deserves the praise and thanks of the
A01 0060 City of Atlanta" for the manner in which the election was conducted.
A01 0070 The September-October term jury had been charged by Fulton
A01 0080 Superior Court Judge Durwood Pye to investigate reports of possible
A01 0090 "irregularities" in the hard-fought primary which was won by
A01 0100 Mayor-nominate Ivan Allen Jr&. "Only a relative handful
A01 0110 of such reports was received", the jury said, "considering the
A01 0120 widespread interest in the election, the number of voters and the size

A01 0130 of this city″. The jury said it did find that many of Georgia's
A01 0140 registration and election laws ″are outmoded or inadequate
A01 0150 and often ambiguous″. It recommended that Fulton legislators
A01 0160 act ″to have these laws studied and revised to the end of modernizing
A01 0170 and improving them″. The grand jury commented on a number
A01 0180 of other topics, among them the Atlanta and Fulton County purchasing
A01 0190 departments which it said ″are well operated and follow generally
A01 0200 accepted practices which inure to the best interest of both governments″.

LOB / WC の例

A01 1 **[001 TEXT A01**]
A01 2 *<*'*7STOP ELECTING LIFE PEERS**'*>
A01 3 *<*4By TREVOR WILLIAMS*>
A01 4 |^A *0MOVE to stop ¥0Mr. Gaitskell from nominating any more Labour
A01 5 life Peers is to be made at a meeting of Labour {0M P}s tomorrow.
A01 6 |^¥0Mr. Michael Foot has put down a resolution on the subject and
A01 7 he is to be backed by ¥0Mr. Will Griffiths, {0M P} for Manchester
A01 8 Exchange. A01 9 |^Though they may gather some Left-wing support, a large majority
A01 10 of Labour {0M P}s are likely to turn down the Foot-Griffiths
A01 11 resolution.
A01 12 *<*7*'ABOLISH LORDS**'*>
A01 13 |^*0¥0Mr. Foot's line will be that as Labour {0M P}s opposed the
A01 14 Government Bill which brought life peers into existence, they should
A01 15 not now put forward nominees.
A01 16 |^He believes that the House of Lords should be abolished and that
A01 17 Labour should not take any steps which would appear to *″prop up**″ an
A01 18 out-dated institution.

Frown / FLOB の例

A01 1 <#FROWN:A01¥><h_><p_>After 35 straight veto victories, intense
A01 2 lobbying fails president with election in offing<p/>
A01 3 <p_>By Elaine S. Povich<p/>
A01 4 <p_>CHICAGO TRIBUNE<p/><h/>
A01 5 <p_>WASHINGTON - Despite intense White House lobbying, Congress has
A01 6 voted to override the veto of a cable television regulation bill,
A01 7 dealing President Bush the first veto defeat of his presidency just
A01 8 four weeks before the election.<p/>
A01 9 <p_>Monday night, the Senate overrode the veto 74-25. the same
A01 10 margin by which the upper house approved the bill last month and
A01 11 comfortably above the two-thirds majority needed.<p/>
A01 12 <p_>Not one senator changed sides, a blow to Bush's prestige after
A01 13 he had heavily lobbied Republican senators, urging them not to
A01 14 embarrass him this close to the election.<p/>

ACE の例

<section> <title>B - PRESS: EDITORIALS</title>
<sample><id> B01</id>

<source> The Australian </source>
<note>2012 words</note>
<subsample>
<id> B01a</id> <source>The Australian - 28 October 1986</source>
<h> Forgive and forget?</h> RE Nigel Jackson's letter (Forgive And Forget, 2/10), if it were not so sad I could have laughed. As a little girl in England, the little Christian children constantly reminded me that I had killed Jesus. This was taught in Sunday school and they didn't love me and they never forgave me. Wasn't I lucky not to have lived in Europe? I might not have been here to read Mr Jackson's plea to love and for+give my enemies as they have failed to do after almost 2000 years. <bl> LILY GILD Yokine, WA </bl>
THE attitude of "forgive and forget" Nazi war criminals overlooks the vital point - jus+tice. If large-scale and systematic atrocities are left unpunished then we are saying that this behaviour is acceptable. It is not the revenge of crimes committed that is the issue, but upholding the moral code of justice which underlies Western societies. <bl> UNA BELL Mundaring, WA </bl> </subsample>
</sample>

Helsinki の例

<B cochroa2>
<Q chrona2 hist nn o2>
<N CHRON A EARLY>
<A X>
<C O2>
<O 850-950>
<M 850-950>
<K CONTEMP>
<D WS>
<V PROSE>
<T HISTORY>
<G X>
<F X>
<W X>
<X X>
<Y X>
<H X>
<u x>
<E X>
<J X>
<I x>
<Z NARR NON-IMAG>
<s sample x>
[^TEXT: CHRONICLE MS A EARLY (O2). TWO OF THE SAXON CHRONICLES PARALLEL, VOL. I. ED. C. PLUMMER. OXFORD: THE CLARENDON PRESS, 1965 (1892). PP.
2.1 - 112.5^] [^THE TEXT BELOW CONTAINS LATER INTERPOLATIONS GIVEN IN ITALICS OR

SMALL PRINT IN THE EDITION.^] [^B17.1^]
<P 2>
<R 1>
+TY GEARE +TE W+AS AGAN FRAM CRISTES ACENnesse cccc wintra. & xciiii uuintra.
+ta Cerdic & Cynric his sunu cuom up +at Cerdices oran. mid v scipum. & se
Cerdic w+as. Elesing. Elesa. Esling. Esla. Gewising. Giwis. Wiging. Wig.
Freawining. Freawine. Fri+tugaring. Fri+tugar. Bronding. Brond B+ald+aging.
B+ald+ag Wodening.
<R 7>
Ond +t+as ymb vi gear +t+as +te hie up cuomon geeodon West Seaxna rice, & +t+at
uu+arun +ta +arestan cyningas +te West Seaxna lond on Wealum geeodon; & he
h+afde +t+at rice xvi gear, & +ta he gefor +ta feng his sunu Cynric to +tam
rice, & heold xvii winter.
<R 11>
+ta he gefor +ta feng Ceol to +tam rice, & heold vi gear.
<R 12>
+ta he gefor +ta feng Ceolwulf to his bro+tur, & he ricsode xvii gear, & hiera
cyn g+a+t to Cerdice.
<R 14>
+ta feng Cynegils Ceolwulfes bro+tursunu to rice, & ricsode xxxi wintra, & he
onfeng +arest fulwihte Wesseaxna cyninga; & +ta feng Cenwalh to, & heold xxxi
wintra, & se Cenwalh w+as Cynegilses sunu; & +ta heold Seaxburg his cuen an gear
+t+at rice +after him.
<R 17>
+ta feng +ascwine to rice, +t+as cyn g+a+t to Cerdice, & heold ii gear.

PPC の例

以下のような、拡張子に psd のついた Parsed File のみ処理ができます。

((CODE <P_I.42>)) ((LATIN (" ") (FW
Recti) (FW diligunt) (FW te) (, .) (" ") (FW In) (FW canticis) (FW sponsa) (FW
ad) (FW sponsum) (, .) (FW $Est) (FW $rectum) (FW grammaticum) (, ,) (FW rectum)
(FW geometricum) (, ,) (FW rectum) (FW theologicum) (, ,) (FW et) (FW sunt) (FW
differencie) (FW totidem) (FW regularum) (, .) (FW De) (FW recto) (FW
theologico) (FW sermo) (FW nobis) (FW est) (, ,) (FW cuius) (FW due) (FW sunt)
(FW regule) (, .) (FW Vna) (FW circa) (FW cordis) (FW direccionem) (, ;) (FW
altera) (FW uersatur) (FW circa) (FW exteriorum) (FW rectificacionem) (, .) ("
") (FW Recti) (FW diligunt) (FW te) (E_S .) (" ")) (ID CMANCRIW,I.42.2)) (
(IP-MAT-SPE (" ") (NP-VOC (N Lauerd)) (, ,) (" ") (IP-MAT-PRN (VBP sei+d)
(NP-SBJ (NPR$ Godes) (N spuse)) (PP (P to) (NP (PRO$ hire) (ADJ deorewur+de) (N
spus)))) (, ,) (" ") (NP-SBJ (D +Teo) (ADJ richte)) (VBP luuie+d) (NP-OB1 (PRO
+te)) (" ")) (ID CMANCRIW,I.42.3)) ((IP-MAT (NP-SBJ (D +Teo) (CP-REL *ICH*-1))
(BEP beo+t) (ADJP (ADJ richte)) (CP-REL-1 (WNP-2 0) (C +te) (IP-SUB (NP-SBJ
T-2) (VBP liuie+d) (PP (P efter) (NP (N riwle))))) (E_S ,)) (ID CMANCRIW,I.42.4))

Index

ACE, 22
British National Corpus, 20
Brown, 21
COCOA, 13, 21
Collocation, 9
  Copy, 9
  Font, 9
  Save, 9
  Sort, 9
Concordance, 8
  Auto View, 8
  Copy, 8
  Font, 8
  Save, 8
  Sort, 8
Exit, 6
File
  Corpus Settings, 4
  Open, 4
  Option, 5
  Save, 4
Fixed Length, 12, 19
FLOB, 22
Frown, 22
Help
  About, 18
  Contents, 18
Helsinki, 13, 23
Keyword Dialog, 14
  Context, 15
  Keyword, 14
  Match, 14
  Regular Expression, 14
KWIC Format
  Key Position, 5
  KWIC Length, 5
LOB, 22
N-gram, 10
  Copy, 10
  Find, 10
  Font, 10
  N-gram, 6
  Save, 10
PPC, 24
SGML, 12, 20
Stop Word List
  Stop Word, 4
TreeTagger, 10
  Copy, 11
  Find, 11
  Font, 11
  Parameter, 5
  Save, 11
  TreeTagger, 5
View, 6
View Settings
  Auto View, 6
  Font, 6
WC, 22
Word Definition
  Ignore Case, 4
  Separator, 4
Wordlist, 7
  Copy, 7
  Find, 7
  Save, 7
  Sort, 7
正規表現, 15