

「狙ったキーワードは逃さない」

Sniper ver1.2.0 使い方

「狙ったキーワードは逃さない」 Sniper（以下本アプリ）は、Web検索の感覚でキーワードを指定し、目的の拡張子のアイテムを探し出してダウンロードするアプリケーションです。

最大の特徴は、キーワードと標的になる拡張子を指定すれば、放っておいても狙ったアイテムをダウンロードできることです。探索の範囲も切り替えることができ、ターゲットURLが絞れない場合でも念入りに探索することができます。

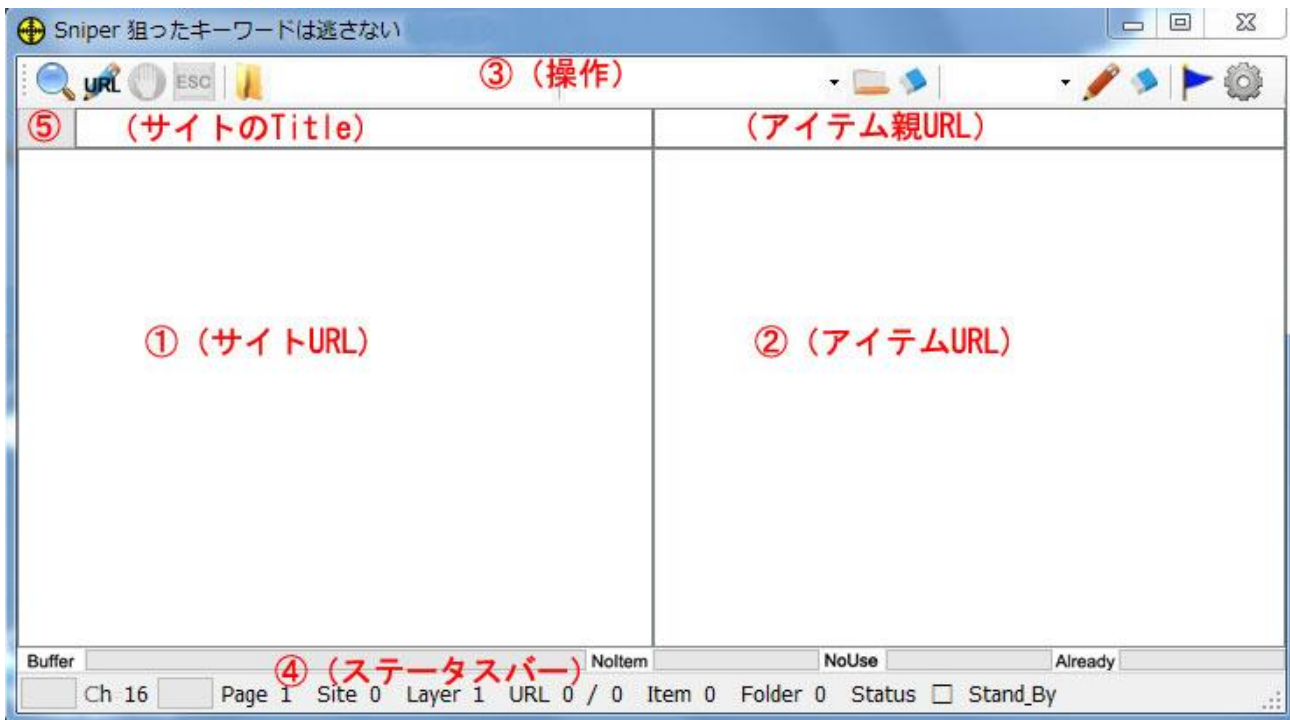
ダウンロードのモードには3つあり、通常の「キーワードモード」に加え、ファイルに列記したキーワードを順次入力して探索する「キーワードFileモード」、取得済みのURLを探索する「バッチURLモード」があります。

ver1. 2.0では、アルゴリズムを見直し、アイテムのヒット率の向上を図りました。、また、ダウンロードを効率よく実行するために、URLを絞り込む機能を充実させました。頻繁に変更しやすいURLフィルタ機能と、恒久的なフィルタ機能とその登録/編集機能とネットの脅威のフィルタ機能です。また、htmlのタイトルに含まれるキーワードでの絞り込み機能等が追加されています。ネットの脅威のフィルタは、ネットの脅威として登録されたURLへのアクセスを回避するだけの簡単なものですが、危うきに近寄らずです、他のフィルタとは切り離して扱うようにしました。お手持ちのウィルス対策アプリから脅威となるサイトのURLが得られればそれを利用してください。

他に、無駄なアクセスを減らすためにアイテム無しが連続したらサイトを放棄したり、URLの重複アクセス回避機や、保存フォルダ形成選択機能などを備えています。

1. 機能概要と画面構成

本アプリの起動画面は次のような画面となります。



アプリの操作画面は大きく4つのブロックに分かれています。

①のURLペインには、アクセスするURL一覧が表示されます。通常モードではWeb検索して得られたURLリストが表示され、バッチURLモードではドラッグ&ドロップされたTextファイルの内容が表示されます。探索するURLをTextファイルに列記しておきここにドロップするのが便利です。

上にある(サイトのTitle)には、ここにリストアップされたURLの内、アクセス中のサイトのTitleが表示されます。これは、アイテムを保存するときの子フォルダ名として用いられます。(キーワードが親フォルダになります。詳しくは、**2.4 アイテムの保存方法**参照)

②のItemURLペインには①にリストアップされたURLにアクセスして、得られたItemURLがリストアップされます。リストアップされるタイミングはダウンロードチャンネルでダウンロードURLを開始するタイミングです。

上にある(アイテム親URL)にはItemURLを得たhtmlのURL(アイテムの親URL)が表示されます。

③の操作部には保存フォルダ、キーワード、ターゲット拡張子、動作ボタン、設定ボタンなどの操作ボタン類が並んでいます。

④のステータスバーには、動作状態を示す種々のステータスが表示されます。

最も左のプログレスバーは動作インジケータで始終動き正常に動作していることを示します。このほか、ダウンロードチャンネル稼働状況やアクセスしているURLの順番、URLの階層、ダウンロードアイテム数、作成フォルダ数などが表示されます。

⑤は、ページ検索間隔を制限するタイマーです。プログラムからWeb検索を実行するとき、サーバーでアクセス頻度を監視する仕組みがあり、連続してアクセスすると、一定時間アクセスが出来なくなります。本アプリではWeb検索を実行する間隔を1分以上とるようにしています。⑤のプログレスバーに緑色の表示が無くなるまで新たなWeb検索が制限されます。

この1分監視は設定でOn/Off出来ます、デフォルトはOnになっています。

2. 先ず使ってみましょう

特別な操作は必要無く、キーワードと拡張子を指定するだけで動かすことができます。まずは、デフォルト設定のままで動かしてみましょう。

本アプリには3つの動作モードがあります、それぞれ順を追って説明します。

(1)**キーワードモード**：キーワードをWeb検索して関連するURLを取得し、そのサイトにある指定の拡張子のアイテムをダウンロードします。

(2)**キーワードFileモード**：キーワードモードは検索語を一つ一つ設定してダウンロードしますが、キーワード設定を自動で行うモードです。このモードを使うと無人運転ができます。

(3)**バッチURLモード**：取得済みのURLをまとめてURLペインにドラッグ&ドロップし、順にアクセスして指定の拡張子のアイテムをダウンロードします。

(2)のキーワードFileモードは設定で選択します。(1)と(3)のを切り替えるためのボタンはありません。。

(1)と(3)の切り替え、はダウンロード開始ボタンをクリックする前に、URLペインにURLのテキスト（ファイル）をドラッグ&ドロップすると、直後のスタートボタンクリックで(3)のバッチURLモードになります。（2回目のクリックはキーワードモードになります）


Shiftキーを押しながら開始ボタンをクリックしても(3)のバッチURLモードで動作します。

2.1 先ず使ってみよう <キーワードモード>

キーワードを指定してWeb検索して関連するURLを取得、そのURLのサイトに含まれる指定の拡張子を持つアイテムをダウンロードするモードです。

○数字番号は操作する順番です。





①ダウンロードしたアイテムを保存するフォルダを指定します。このフォルダは保存の**祖父母フォルダ**となります。 ボタンでフォルダブラウザを開くか、フォルダをドラッグ&ドロップするか、フォルダパスURLを直接入力します。フォルダパスはドライブ名を含むフルパスを記入します。(例： D:¥download)

②ターゲットになるキーワードを入力します。この欄をマウスでクリックしてからキーワードを入力し**リターンキー**を押します。リターンキーを押さないと入力したキーワードは有効になりません。入力したキーワードは20個まで記憶されます、▼でプルダウンメニューを開き以前使用したキーワードをクリックして入力できます。


キーワードを名前に持つフォルダが祖父母フォルダの下に作られ保存の**親フォルダ**となります。


フォルダマークのボタンをクリックすると親フォルダがエクスプローラーで開かれます。

隣の消去ボタン  をクリックすると表示されている拡張子が削除されます。

③記入ボタン  をクリックすると入力画面が開きます。ここではダウンロードターゲットのファイル拡張子を入力します。

④拡張子にはピリオドはつけません。カンマ (,) で区切って複数の拡張子が指定できます。欄右端の▼をクリックしてプルダウンメニューを開きクリックして選ぶこともできます。ここで指定した拡張子のアイテムを探索して見つければダウンロードします。

隣の消去ボタン  をクリックすると表示されている拡張子が削除されます。

⑤開始ボタン  をクリックするとURLペインの表示をクリアして探索を開始します。

デフォルトでWeb検索は、1ページ当たり20件で、2ページまで、各URLの第2階層のhtmlに記述されたItemURLまで探索する設定になっています。この条件は変更できます。

※第1階層はURLペインに登録されたURLの階層です。

⑥一連の操作との連続性はありませんが、ここで説明しておきます。

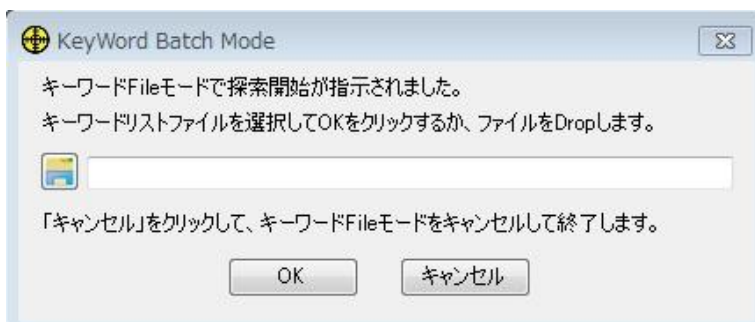
本アプリは、URLに対するフィルタ機能を備えており、探索時に不要なURLを排除できます。このボタンは、フィルタに語句を設定するためのものです。語句の設定については、4. URLフィルタの語句設定を参照して下さい。

フィルタに条件が設定されている場合、⑤の開始ボタンをクリックした後に設定されているフィルタの内容が表示され、その条件でのフィルタリングの可否を聞いてきます。選択肢は3つ、フィルタリング適用、フィルタリング無視、探索中止の中から選びます。

2.2 先ず使ってみよう <キーワードFileモード>

このキーワードFileモードにするには、5. **本アプリの動作条件の設定**の画面最上段にある、「通常探索 ☐ 探索開始時・・・」のチェックをOnにします。

2.1のキーワードモードの操作と異なるのは、②のキーワード指定を飛ばすところです。キーワード指定をせずに④の開始ボタンをクリックすると、次の画面が開きます。



この画面は、キーワードが列記されたキーワードリストファイルを指定するためのものです。

方法は、KeyWordバッチファイルを直接この画面にドラッグ&ドロップするか、ファイルボタンをクリックしてファイルブラウザでバッチファイルを指定してOKボタンをクリックします。直接記入してOKボタンをクリックする方法もあります。

キャンセルをクリックすると、そこで終了します。

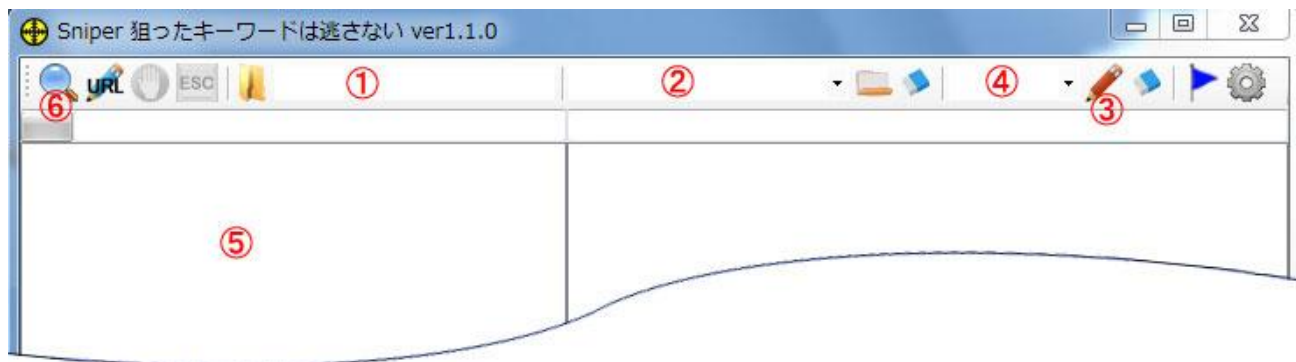
キーワードリストファイルは、一般的なテキストファイル（拡張子.txt）です。書式は、1行に1キーワードで行数制限はありません。空行（リターンのみ）には、『リスト終了』の意味付けがされています。複数行の途中に空行を含めると、その空行のところで動作を止めることができます。

※キーワードモードと同じに、URLフィルタ条件が設定されているとフィルタリングの可否が尋ねられます。

2.3 先ず使ってみよう <バッチURLモード>


アイテムをダウンロードするサイトのURLを直接指定してダウンロードするモードです。ダウンロードするURLを予め取得しているときに使います。

URLペイン⑤にダウンロードするURLを列記したテキストファイルをドラッグ&ドロップします。或いはメモ帳などでURLのテキストを反転選択しURLペインにドラッグ&ドロップします。



①～④は、キーワードモードの①～④と同じように操作して、（祖父母）フォルダ、キーワード（親フォルダ）、拡張子を設定します。この場合キーワードは、キーワードとしての意味はなく、アイテム保存の親フォルダ名となります。

⑤アクセスする**URLのテキストファイルをURLペインにドラッグ&ドロップします**。URLペインにドラッグ&ドロップするとURLペインの表示はクリアされ、ドラッグ&ドロップされた内容が表示されます。

⑥  ボタンをクリックするとURLペインのURL表示はそのままに探索を開始します。ここで注意が必要なのは、URLペインにドロップした直後に開始ボタンをクリックしないとバッチURLモードにならないことです。

不安がある時には、シフトキーを押しながら開始ボタンをクリックすればバッチURLモードで動作します。

※キーワードモードと同じに、URLフィルタ条件が設定されているとフィルタリングの可否が尋ねられます。

キーワードモードとの違いは、スタートする前にURLテキストをドラッグ&ドロップすることです。

2.4 アイテムの保存

ダウンロードされるアイテムは、祖父母フォルダ(2.①欄) の下の 親フォルダ(2.②欄) 内に一括保存されます。

設定により、URLペインの各URL固有のフォルダに**分散保存**することができます。このとき、URLのHTMLの中に書かれたTitleの一部をフォルダ名に使います。このフォルダは**子フォルダ**に相当します。子フォルダは、

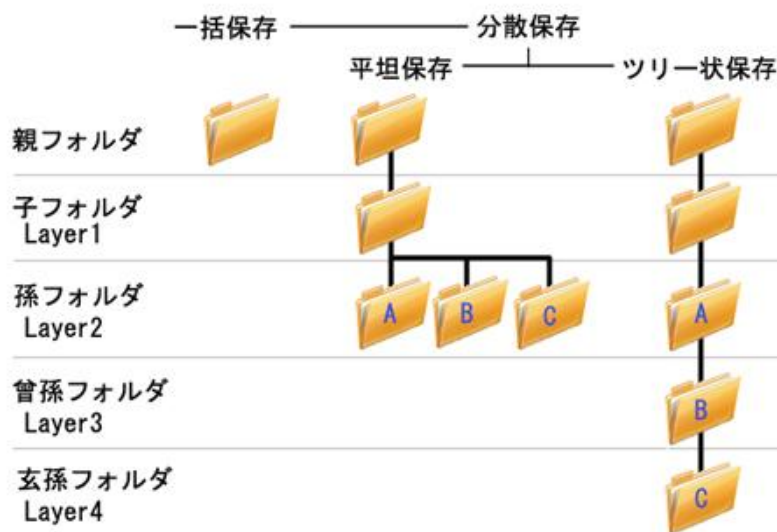
URLペインに書かれた第1階層のURLに対応して形成され、この階層でダウンロードされるアイテムはこの子フォルダに保存されます。

得られたhtmlにTitleが設定されていないケースもあります、そのときには「No Title」がフォルダ名として用いられます。

更に第2階層以降のURLに対しては階層に準じてフォルダをツリー状の形成するか、階層に関係なく平坦に形成するかを選択ができます。

。ツリー状に形成する場合、新たに作られるフォルダは孫フォルダ、曾孫、玄孫・・・と階層に応じて深い位置に形成されます。平坦に保存する場合は階層に関係なく孫フォルダのレベルに平坦に形成されます。

これらを設定の中で選択します。デフォルトでは分散・平坦保存に設定されています。



2.5 右クリック

URLペインを右クリックすると次のようなアプレットが開きます。それぞれの働きは以下の通りです。

	このURLをブラウザで開く
	このURLを探索しないサイトに登録
	このURLをクリップボードにコピー
	クリップボードのURLをListに追加
	語句を検索する
	選択されたURLを削除する
	すべてのURLを削除する
	ネット脅威のURLを追記する

1. このURLをブラウザで開く

URLを選択しておいてここを選択すると、デフォルトのブラウザでそのサイトにアクセスします。

2. このURLを探索しないサイトに登録

探索しないURLをURLペインから拾う機能です。

3. このURLをクリップボードにコピー

反転したURLをコピーします。

4. クリップボードのURLをListに追加

ブラウザなどからコピーしたURLを追加します

5. 語句を検索する

表示されているURLを検索します。

6. 選択されたURLを削除する

表示されているURLリストから削除します。

7. 全てのURLを削除する

表示されているすべてのURLを削除します。

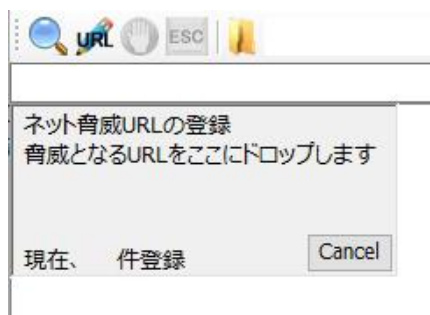
8. ネット脅威のURLを追記登録する

ネット脅威のURLリストが、本アプリが格納されているフォルダに **Threat.txt** の名前で保存されています。初期状態ではアプリ作者が本アプリのデバッグ期間中にウィルス対策ソフトが拾った脅威サイトが記載されています。

このリストに新たな脅威サイトを追記登録することが出来ます。お使いのウィルス対策アプリから脅威サイトのURLが得られたらそれを追加登録すると安心です。登録されるURLは脅威サイトURLの第一階層（ルートURL）の文字列です。この文字列を含むURLのサイトには以降アクセスしなくなります。アクセスを回避したときにはツールバーの旗が数秒間赤色に変わります。

メニューをクリックすると、次の画面のようにドロップのための小窓が開きます。ここに、脅威のURLが記述されたテキストファイル（xxx.txt）をドロップするか、テキストエディタで反転表示させたURL（リスト）をドラッグ&ドロップして登録します。登録されると自動的に小窓は閉じます。

編集機能は備えていませんので、お手持ちのテキストエディタで **Treat.txt** を編集してください。



3. 機能説明

以下に本アプリの操作部の機能を説明します。説明の丸番号は手順ではありません。

3.1 画面左翼操作ボタン



①探索開始ボタンです。

キーワードモード、キーワードFileモード、バッチURLモードのダウンロードの開始ボタンです。

②URLフィルタの単語設定

本アプリには2系統のURLフィルタが備わっています。ここでフィルタに使う単語/文字列を指定します。(URLフィルタは4節を参照してください。)

③探索停止ボタン

すべてのダウンロードを停止します。ダウンロードを開始するとボタンが有効となり、赤背景の配色となります。

キーワードFileモードで動作中は、機能の選択が求められます。このモードでは停止機能をどのように働かせるかを選択します。現在のキーワードの探索終了を待って停止するか、直ちに停止するかを選択します。

④探索中URLEスケープボタン

URLペインのダウンロード中のURLの探索動作を中断して抜け、次のURLの探索動作を始めます。ダウンロードを開始留守とボタンが有効となります。

キーワードFileモードで動作中は、機能の選択が求められます。このモードではエスケープ機能をどのように働かせるかを選択します。現在探索中のキーワードを直ちに抜け次のキーワードに進むか、(URLペイン上で)現在のURLを直ちに抜け次のURLに進むかを選択します。

⑤フォルダブラウザを開きます。

フォルダブラウザを開いてダウンロードアイテムを保存する祖父母フォルダを選択・指定します。

⑥ダウンロードしたアイテムを保存するフォルダを指定します。

ここで指定するフォルダがダウンロードアイテムを保存する**祖父母フォルダ**となります。(更に、キーワードもフォルダ名として使われ、これが保存の親フォルダとなります。Web検索で取得したURLからダウンロードするアイテムはこの親フォルダ下に保存されます。)

エクスプローラーなどからフォルダをここにドラッグ&ドロップして指定することもできます。

3.2 画面右翼操作ボタン



①キーワードの指定

本アプリの肝となるキーワードを入力します。キーワードを入力するときは、①のエリアをクリックして反転表示状態にしてからキーボードで入力し、リターンキーを押します。入力したキーワードは20件まで記憶され再利用に供されます。

キーワードプルダウンメニューボタンをクリックすると、最大20件のキーワードがプルダウンメニューに表示され、クリックしてそのキーワードが入力できます。派生キーワード入力に便利です。

選択したキーワードを修正して使用する場合も最後に「リターンキー」を押してください。

※キーワード入力後リターンキーを押さないと、そのキーワードは保留状態になっています。このままで開始ボタンをクリックすると保留は解除されます。また、保留状態のまま新たなキーワードを入力すると保留されていたキーワードは廃棄されます。リターンキーを押すことで20件のドロップメニューに登録されます。

②アイテム保存フォルダ展開ボタン

現在設定されている(祖父母フォルダ+親フォルダ)をエクスプローラーで開きます。

③キーワード削除ボタン

現在表示されているキーワードを削除します。 Shiftキーを押しながらクリックすると全項目が削除されます。

④ターゲット拡張子の指定

ダウンロードするアイテムの拡張子をプルダウンメニューから選択します。この欄に直接入力できません。

⑤ターゲット拡張子登録ボタン

このボタンをクリックすると拡張子入力画面が開きます。

アイテムの拡張子をピリオド (.) を付けずに登録します。拡張子をカンマ (,) で区切り、複数のアイテムを同時に登録することができます。

⑥ターゲット拡張子削除ボタン

プルダウンメニューに登録されている拡張子を削除するボタンです。削除したい拡張子を選んでからこのボタンをクリックします。

Shiftキーを押しながらクリックすると全項目が削除されます。

⑦終了通知フラッグ/脅威サイト忌避フラッグ

この旗が青色の時、ダウンロード動作が終了すると通知してくれます。白抜きの時には動作が終了しても通知はありません。

脅威サイトリスト (Threat.txt)に脅威サイトのルートURLを登録してあるとき、該当するサイトのアクセスが回避されますが、会費動作をした時にこのフラッグが数秒間赤色に替わります。

⑧動作条件設定ボタン

本アプリの動作条件を設定します。条件設定は5節を参照してください。

3.3 URLの抽出

本アプリでは次のような基準でURLを抽出しています。

(1)URLペインのサイトURL

URLの抽出はページの内容を入れ、ページあたりの件数を設定します。

単語を検索して設定件数のURLをWebから取得したらそのURLをURLペインに登録します。このとき、設定5.1の⑫「探索しないURL」に登録された文字列を含まないURLが登録されます。

ページのURLの探索が終了したら次のページを読み込みますが、現在のページが得られた時から1分以上経過していないと次のページを取得できないようになっています。1分に満たない場合、「サイトのTitle欄」と「アイテム親URL欄」の背景色が交互に変わりその旨表示されます。

(2)ItemURL

指定されたいずれかの拡張子を含むURLで、

html上で `<a xxxxx href="ItemURL. 拡張子?yyyyy" zzzzz>` のように記述されているURLを抽出します。

xxxxxx、?yyyyy、zzzzzの部分がない様式にも対応します。直貼りのときは、`<a xxxxx href が <img src` になります。

(3)第2階層を含み下層へのLinkURL

URLフィルタを適用するときには、登録された単語を含み、且つRootURLを含むURLで、

URLフィルタを適用しないときには、サイトURLで決まるRootURLを含むURLで

html上で `<a xxxxx href="LinkURL" zzzzz>`

または、 `<ahref="LinkURL">` のように記述されているURLを抽出します。

(4) 下層のLinkURLの並べ替え

下位階層のLinkURLは、html上に並んでいる順番で取得されます。本アプリでは、効率化のためにアイテムのないリンクや外部サイトのURLなどをサクセス対象から外す機能があり、外されたURLが指定した数に達するとそのサイトを放棄する(On/Off可能)仕組みになっています。

無効なURLが先頭から並んでいると、目的のURLに到達する前にサイトを放棄してしまう可能性が高くなります。

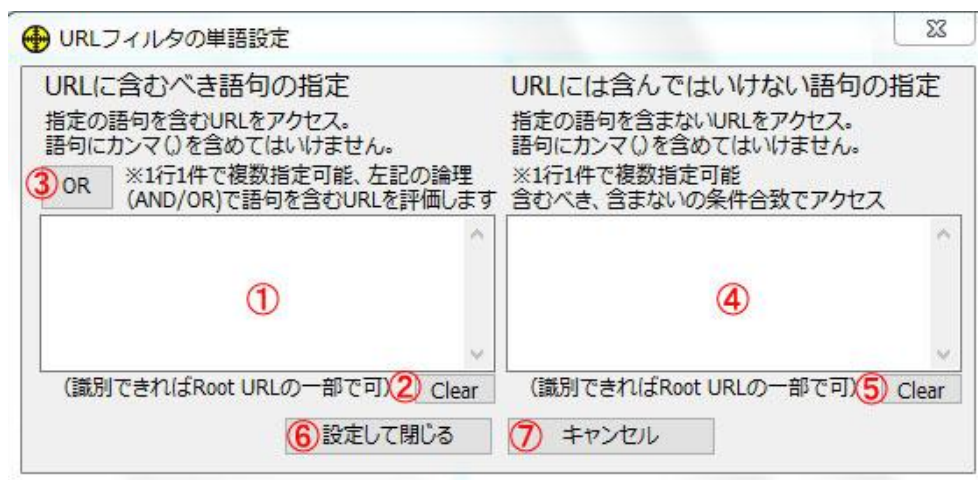
経験的に望むリンクは文字数が多い傾向があります。そこで本アプリではLinkURLを取得すると文字数の多い順に並べ替え早い時期に所望のURLをアクセスできるようにして取りこぼしの機会を減らしながら効率化を図っています。

4. URLフィルタの語句設定

Web探索時のURLフィルタの語句登録を行います。ここで設定するURLフィルタの条件に合致しないURLにはアクセスしません。

フィルタは含むべき語句と含んではいけない語句2系統を設け、両方のフィルタを通過するURLのみにアクセスします。

語句は複数同時に指定することができ、含むべき語句の場合、すべての語句を含む（AND）か、いずれかの一つを含む（OR）かを切り替え指定することができます。そのロジックはボタンで切り替えます。



フィルタの条件を論理的に表すと

(①語句 1 AND/OR ①語句 2) AND (④語句 1 OR ④j語句 2)

この2系統のフィルタを同時に満たすURLのみにアクセスします。①、④が空白のときにはそれぞれのフィルタ機能は働きません。

探索を開始するとき、設定されているフィルタの内容を表示してフィルタリングの可否が尋ねられます。選択肢は3つで、フィルタリングする、フィルタリングしない、探索中止から選びます。フィルタの単語、文字列が設定されていないときには何も尋ねられません。

①、④URLに含まれるべき語句/含んではいけない語句の入力

URLに含まれるべき語句、URLに含んではいけない語句を1行に1件登録します。複数件登録できます。

語句は手入力あるいは、クリップボードを介して貼り付けてください。

②、⑤単語消去ボタン

①、④に登録されている語句をすべて消去します。

③URLに含まれる語句の評価ロジック

①に登録されている複数の語句の評価方法を指定します。ANDはすべてを同時に含むとき、ORはどれか一つ含むとき、そのURLにアクセスします。

⑥設定した条件を登録して閉じる

設定した条件をアプリに登録して設定ウィンドを閉じます。

⑦キャンセル

設定画面の変更はすべてキャンセルして設定ウィンドを閉じます。

※含んではいけない語句に関しては設定の5.1の⑭でも指定することが出来ます。5.1⑭の方は単独で働くもので、恒久的に排除するURLを指定する場合に有効で、例えば「人名」をキーワードにすると、twitterやwikipediaが必ず対象URLに含まれます。このようなURLを恒久的にアクセスしない時に5.1の⑭を用いることを想定しています。

5. 本アプリの動作条件の設定

ここでは本アプリの動作条件を設定します。Web検索に関する設定、URL重複アクセスに関する設定、URLのHTMLに関する設定、保存フォルダの設定などが行えます。順次ブロックごとに説明します。

5.1 動作条件設定と非サクセスサイトの登録

Sniper条件設定

①・KeyWordモード or バッチURLモードで探索 ☐ KeyWordFileモードで探索

②③検索ページ数とLayer深度(URLペインが1) 2 ページ 2 階層

④・検索1ページ当たりの取得URL数 ☐ 10 ☒ 20 ☐ 50

⑤・同時にダウンロードするチャンネル数 ☐ 4 ☒ 8 ☐ 16

⑥・Item無しURLが指定数になるまでサイト維持 ☒ 10 % 10 回で放棄

⑦・既読URLが続いてもアクセスを継続 ☒ 10 回でサイトを放棄

⑧・第2階層以降のURLの最短文字数指定 ☐ 30 文字以上に制限

⑨・ダウンロードアイテムが少なくてもサイト継続 ☐ 5 未滿連続で放棄

⑩・Web Titleをフォルダ名に利用、最大文字数は 40 文字に制限する

⑪・Web検索の間隔に制限を設けない ☐ 最少1分間隔に制限する

⑫・ページ直貼りアイテム(<... src=)の扱い ☐ 直貼りもダウンロードする

⑬・過去アクセス済みURLとの重複を許容する ☒ 過去分との重複を回避する

⑭・先頭ItemURLだけ拡張子偽装チェック&解除 ☐ 全ItemURLをチェック&解除

⑮・探索から外すサイトの指定 ⑯確認・編集 ☐ これらも同等に探索する

・URLペインのURLの階層下を探索範囲とする ☒ 同サイト上位階層も探索

・URLペインのURLのサイト内のみを探索する ☐ サイト外のURLも探索する

・Layer2以降末尾がhtml()のURLのみアクセス ☒ すべてのURLを探索する

・html取得に際してのサイトアクセス時間制限 10 秒でアクセス中止

・ItemをKeyWord名親フォルダに一括保存する ☒ title名フォルダに分散保存

・Title名フォルダを階層化しツリー状に形成する ☒ 階層に関係なく平坦に形成

・html TitleにKeyWord第一語句あればアクセス 2 階層迄適用(0でOff)

・アイテムをダウンロードしたURLの扱い ☒ 対応フォルダにURLを保存

・Itemダウンロード時間制限(時間超過で放棄) 1 分(全Item共通)

・フォントの設定 MS UI Gothic 9.75

・アイテムペインのURLをダブルクリックして画像を表示する時に用いるピクチャー

重複Table消去 設定して閉じる 出荷状態に戻す キャンセル

①探索モード選択

開始ボタンをクリックした時の動作モードを設定します。通常キーワードモードで探索を開始します。Shiftキーを同時押しすると、バッチURLモードで探索を開始します。

ここにチェックを入れると探索開始時、キーワードリストファイルが求められます。キーワードリストファイルをドラッグ&ドロップするか、ファイルブラウザを使ってファイルを選択します。

②キーワード検索実行ページ数

キーワードの検索を何ページ実行するかをUp/Downボタンで設定します。

ページ概念は、通常のWeb検索で1回に表示される検索結果と同じです。デフォルトで「2」が設定されています。

③URL探索のLayer深度

Web検索で取得したURLをどの階層まで掘り下げて探索するかを指定します。第1LayerはURLペインにリストアップされるURLの階層になります。その子供が第2Layerとなります。

第2Layerを設定した場合、その階層のURLに貼り付けられたItemURLを取得しますが、これ以上階層を潜らないのでLinkURLは取得しません。

深くするほどアクセスするURLが増えるため、全体としてダウンロードの進み具合が遅くなります。デフォルトでは「2」に設定されています。

④キーワード検索時の1ページ当たりのURL取得件数

Web検索で取得する1ページあたりのURL件数を指定します。10、20、50件から選択出来ます。デフォルトで「20」に設定されています。

⑤同時にダウンロードするチャンネル数

本アプリのダウンロード速度を決める設定で最大で同時16チャンネルまで対応しています。設定は4、8、16チャンネルから選択します。デフォルトは「8」です。

⑥Item無しのURLが連続するときの対応（1．2から回数指定に%指定を追加）

Itemの無いURLが続くときには、そのサイトではアイテムが取得出来ない可能性が高くなります。効率化のためこのような時にそのサイトを放棄して次のサイトに移行することが出来ます。

この機能をOnにして、Item無しの連続を許容する回数を指定します。指定には、回数指定と、%指定の2つの指定があります。%指定は、サイトの第1階層で得られるLinkURL総数の何%の値を許容回数とするかを決めます。回数は直接回数を指定します。この%指定と回数指定の数値の小さいほうが許容回数として設定されます。また、許容回数の最小値は5回に制限されています。

Item無しURLは、「NoItems」としてカウントされます。「NoItems」が許容回数を超えたとき、そのサイトは放棄して次のサイトに移行します。途中、一つでもItemがあった場合には、カウントはリセットされます。「NoItems」の指定値に対する進捗はステータスの⑫のプログレスバーに表示されます。

デフォルトは10%、20回になっています。（200個以上のURLは20回まで許容、それ以下は10%の回数まで許容）

※1 設定要領：大きなサイトだとLinkURLが数百と多く、時には1,000も越えます。URLの多少で許容回数のバラつきも大きくなりバランスが取りにくくなりますが、LinkURLの多いサイトに対応し回数指定で最大値を決め、%指定でLinkURLの少ないサイトに対応すると比較的バランスがとれるようになります。

※2 5.2の②、③、④の条件でアクセスされないURLのカウントにもここで設定した許容回数が用いられます。

⑦既読URLが連続するときの対応

既読URLが連続するときは、既にアクセス済みのサイトである可能性が高くなります。効率化のためこのような時にそのサイトを放棄し次のサイトに移行することが出来ます。

この機能をOnにして、既読の連続を許容する回数を指定します。その回数を超えたとき、そのサイトを放棄して次のサイトに移行します。途中で一つでも既読でないURLがあった場合、回数のカウントはリセットされます。

なお、既読URLの状態は後述する6. ステータスの⑬のプログレスバーに表示されます。

⑧第2階層以降のURLの文字数制限

URLペインに第1階層のURLが表示され、そのURLを順にアクセスしアイテムを探します。階層指定が2以上の場合は、第2階層のURLを取得します。この時に文字数が指定文字数以上のURLのみを取得するための機能です。

経験的に第2階層以降のURLは、文字数がある程度長いものの方がアイテムのヒット率が高い傾向があります。短いURLを切り捨てることでWebアクセスを減らし、ダウンロードの効率を向上させることができます。但し、短いURLでヒットすることもあり、設定値によっては取りこぼすこともあります。

⑨ダウンロードしたアイテムが少ないのが連続したらサイトを放棄

一つのURLに対してダウンロードしたアイテムが指定個数より少ないことが一定回数に達した場合、そのサイトを放棄する機能です。ここで、URLあたりの最低アイテム数を設定します。上限回数は⑥のアイテムなしの上限回数をそのまま利用します。

⑩WebのTitleをフォルダ名に使用するときの文字数制限

サイトのhtmlのTitleをアイテムを保存するフォルダに使用しますが、Titleをフォルダ名にする時の文字数を制限します。

⑪We検索間隔に制限を設けない

ページ検索する時間間隔を制限するタイマー機能があります。このタイマーをOffにすると制限が無くなりますが、アクセス頻度が高くなるとサーバーから制限を受けることがあります。

設定でこのタイマーをOn/Offすることができます。デフォルトはOnになっています。

⑫ページ直貼りイメージ(img src)の扱い

ダウンロードするアイテムは直接リンクのアイテムが基本ですが、直貼りの画像アイテムにも対応しています。

直接リンクはHTMLタグが `` 表記のアイテムです

直貼りの場合HTMLタグが `` 表記のアイテムです。

直貼りイメージもダウンロードするときにはここにチェックします。デフォルトは「チェック無し」です。

⑬過去アクセス済みURLとの重複を許容する

本アプリではアクセスしたURLの記録を残し重複を回避することができます。記録は消去するまで保存されます。

デフォルトでは「チェック有り」です。

重複回避するとき、すべてのURLの記録を残すと膨大な量になってきます。アクセス速度にも影響が出ますので時々記録を消去することをお勧めします。（重複Table消去参照）

なお、アクセスの記録は本アプリの実行ファイルがあるフォルダに Hush.dat の名前で作成されます。

⑭末尾が拡張子の偽装URLの対応

サイトによっては、ItemURLの記述であっても実際にアイテムは無く通常のLinkURLということがあります。

この設定は、アイテムに偽装されたURLをLinkURLとして扱うための設定です。デフォルトはOffになっています。

本アプリでは、ItemURLを取得したときに、そのひとかたまりの先頭のURLをLinkURLのようにアクセスしてその応答でItemかリンクかの選別をしています。この動作は設定のOn/Offに関わらず実行されます。その

URLがLinkURLであった場合には、ひとかたまりのURLはすべてLinkURLとして扱うようになっています。この場合、先頭がItemURLでもその次から偽装URLがあってもそれは拾えないことになります。

この設定は、ItemURLを先頭の1URLだけでなく、ひとかたまりのURLすべてをチェックするように指定するものです。取りこぼしは無くなりますが毎回Webにアクセスが入るため速度が遅くなります。

偽装URLのアイテムのダウンロードでは、主画面④ステータスの「Status □ xxxx」の表記の□が■に変化します。

⑮探索しないサイトの指定

キーワード検索をすると必ずと言っていいほどに現れる常連サイトがあります。そのようなサイトはアクセスせずスルーして探索時間を短縮することができます。

この機能は、「4. URLフィルタ語句設定」でも提供されていますが、こちらは恒久的にアクセスしないURLをスルーすることを想定しています、4.の方は、頻繁に設定を変えることを想定しています。

探索しないサイトはあらかじめ登録しておきます。

「これらも同等に探索する」にチェックを入れると登録された文字列を含むURLもアクセスされます。デフォルトは「チェック無し」です。

⑯探索しないサイトの一覧と編集

「確認・編集」ボタンをクリックすると「非サクセスサイトの登録」に説明する画面が開き、登録されている文字列が昇順ソートされ表示されます。

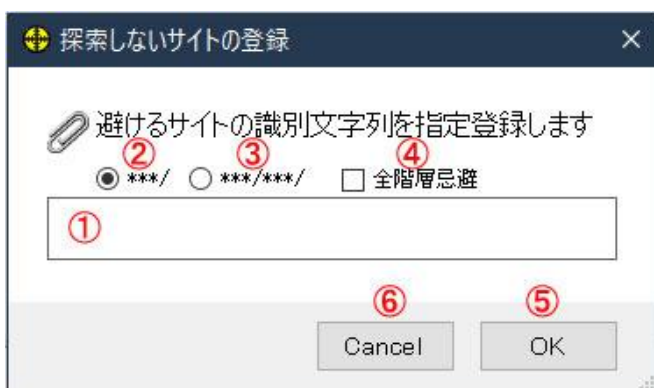
ここでサイトのURLを判定する文字列の編集、確認を行います。（手入力になりますが登録もできます。）

⑮の設定で、ここに登録された文字列を含むURLにはアクセスし無くなります。

○非アクセスサイトの登録

非サクセスサイトの登録は、サイトURLペインに表示されるURLを右クリックして「このURLを探索しないサイトに登録」をクリックして行るか、アイテムURLペインにサイトのショートカットをドロップして行います。登録では下記画面が開きます。

ダウンロード動作中であっても、若干動きが鈍くなりますが登録できます。



本アプリでは/で区切られたURLの文字列を処理単位としています。各部の機能は以下の通りです。

①登録される文字列が表示されるテキストボックス

右クリックしてURL、ドロップしたURLから第1階層文字列、第2階層文字列が抽出表示されます。このボックス内で編集することもできます。

②URLの第1階層文字列を抽出する

URLのhttp://www. の部分は削除して特徴的な部分だけを抽出します。/ も一緒に抽出します。

③URLの第1階層と第2階層の文字列をまとめて抽出する

②に合わせて第2階層分を抽出します。/ も一緒に抽出します。

④適用する階層を指定する

全階層で適用するか、第2階層以降で適用するかを決めます。チェックをOnにすると全階層で適用、Offで第2階層以降での適用になります。

⑤テキストボックス①の文字列を登録する

登録された内容は設定画面を閉じると有効になります。

⑥操作をキャンセルする

○非アクセスサイトの編集

⑩ボタンをクリックすると次の画面が開きます。この画面で登録されている非アクセスサイトの条件を編集することができます。

非アクセスサイトはフルパスでも問題ありませんが、本アプリでは、サイトを識別できる最短文字列を用いています。URLの/で区切られる、第1階層と第2階層の文字列で非アクセスサイトを指定しています。

編集画面は、2つのテキストボックスが設けられており、記述された条件が適用される階層が異なります。上のテキストボックスは全階層で適用され、下のテキストボックスは第2階層以降で適用されます。各部の機能は以下の通りです。



①「1. 全階層で忌避するURL」 の文字列は、アクセスする階層に関係なくURLの判定に用いられます。

②「2. 第一階層は許可、第二階層以降は忌避するURL」 の文字列は、第一階層の判定には用いません。

ダウンロード効率向上のための機能です。①と②はわずかな違いですが、第1階層に有効なアイテムがあるけど第2階層以降は不要と言うようなケースが多々あったので設けた機能で、全階層を非アクセスとすると都合が悪くなることの救済措置になります。

③編集結果を登録

登録された内容は設定画面を閉じると有効になります。

④操作をキャンセルする、編集内容は破棄される

⑤テキストボックス1, 2間に同じ文字列があったら反転表示する

上下のテキストボックスで重複している場合、②の第2階層対応の文字列が無意味になります。上のテキストボックスが全体を網羅するためです。そのような不必要な文字列を削除するための機能です。

⑥すべての記述に対して、従属関係にある文字列を反転表示

URLの階層で従属関係にある文字列が混在すると、文字列を登録した意図が損なわれます。同じテキストボックス内、上下のテキストボックス間で従属関係にある文字列を探し、不必要な文字列を削除するための機能です。

⑦テキストの反転表示を切り替える

重複文字列や従属関係の文字列を同時に反転表示すればいいのですが、能力的に同時に反転表示できないため切り替えて表示しています。反転個所がボックス1, 2間で交互に替わるか、どちらか一方で反転個所が切り替わります。反転表示されている文字列が削除の対象になります。

⑧反転表示している文字列を削除する

削除する文字列を反転表示させておいて、このボタンをクリックして削除します。

○文字列の従属関係について

(A) google/ と (B) google/text/ という2つの文字列で従属を説明します。URLとしては、(A)が (B) の上位階層になり、(A)が (B)を包含しています。

(A)と (B)が混在する場合、(B)は (A)に包含されているため指定する意味は無くなります。

ボックス①に (A)が、ボックス②に (B)が記述されているとき、全階層で google/ の文字列を含むURLはアクセスされなくなります。google/text/ は google/ に包含されているので有ってもなくても一緒になります。

しかし、①に (B)、②に (A)が記述された時には (B)も仕事をします。第一階層で googl/text/ を含むURLを忌避し、第2階層以降では (A)の googl/ が働き google/ を含むURLを忌避します。第一階層だけに異なる動きを設定できます。

⑥の「従属」はこのような従属関係になる文字列を探しだすためのものです。

5.2 動作条件設定

Sniper条件設定

- KeyWordモード or パッチURLモードで探索 ☐ KeyWordFileモードで探索
- 検索ページ数とLayer深度(URLペインが1) 2 ページ 2 階層
- 検索1ページ当たりの取得URL数 ☐ 10 ☒ 20 ☐ 50
- 同時にダウンロードするチャンネル数 ☐ 4 ☒ 8 ☐ 16
- Item無しURLが指定数になるまでサイト維持 ☒ 10 % 10 回で放棄
- 既読URLが続いてもアクセスを継続 ☒ 10 回でサイトを放棄
- 第2階層以降のURLの最短文字数指定 ☐ 30 文字以上に制限
- ダウンロードアイテムが少なくてもサイト継続 ☐ 5 未満連続で放棄
- Web Titleをフォルダ名に利用、最大文字数は 40 文字に制限する
- Web検索の間隔に制限を設けない ☐ 最少1分間隔に制限する
- ページ直貼りアイテム<... src= >の扱い ☐ 直貼りもダウンロードする
- 過去アクセス済みURLとの重複を許容する ☒ 過去分との重複を回避する
- 先頭ItemURLだけ拡張子偽装チェック&解除 ☐ 全ItemURLをチェック&解除
- 探索から外すサイトの指定 ☐ これらも同等に探索する
- ① • URLペインのURLの階層下を探索範囲とする ☒ 同サイト上位階層も探索
- ② • URLペインのURLのサイト内のみを探索する ☐ サイト外のURLも探索する
- ③ • Layer 2以降末尾がhtml()のURLのみアクセス ☒ すべてのURLを探索する
- ④ • html取得に際してのサイトアクセス時間制限 10 秒でアクセス中止
- ⑤ • ItemをKeyWord名親フォルダに一括保存する ☒ title名フォルダに分散保存
- ⑥ • Title名フォルダを階層化しツリー状に形成する ☒ 階層に関係なく平坦に形成
- ⑦ • html TitleにKeyWord第一語句あればアクセス 2 階層迄適用(0でOff)
- ⑧ • アイテムをダウンロードしたURLの扱い ☒ 対応フォルダにURLを保存
- ⑨ • Itemダウンロード時間制限(時間超過で放棄) 1 分(全Item共通)
- フォントの設定 MS UI Gothic 9.75
- アイテムペインのURLをダブルクリックして画像を表示する時に用いるビューアー

①URLペインのURLの階層を探索範囲とする

本アプリではRootURLを設定して探索の範囲を規制しています。

通常、http://www.abc/def/ghj/klm.html のURLに対して RootURL は http://www.abc/def/ghj として、抽出されたURLの内、RootURLを含むURLにアクセスし探索範囲を絞ります。

「同サイト上位階層も探索」にチェックした場合、http://www.abcをRootURLとして、抽出されたURLの内、RootURLを含むURLにアクセスします。結果、探索範囲は広がります。

デフォルトは「チェック有り」です。

②URLペインのURLのサイト内のみ探索する

http://www.abc/def/ghj/klm.html のURLに対して 設定されるRootURLを含むURLにアクセスするのが基本です。

「サイト外のURLも探索する」にチェックした場合、RootURLを含まないURLにもアクセスするようになります。結果、探索範囲が大幅に広がります。

デフォルトは「チェック無し」です。

③Layer2以降末尾がhtmlのURLのみを探索

URLペインのURLにアクセスして得られるURLに対する処理で、末尾のhtmlの有無をみて動作を決めます。

「すべてのURLを探索する」にチェックした場合、URLの末尾のhtml有無に関わらずすべてのURLにアクセスするようになります。結果、探索範囲が広がります。

デフォルトは「チェック有り」です。

※以上②、③、④の条件から外れてアクセスされなかったURLを「NoUse」としてカウントしています。このカウント数が5.1の⑥で指定した数の2倍に達するとそのサイトは放棄されます。「NoUse」の指定値に対する進捗はステータスの同名のプログレスバーに表示されます。

④各URLの取得に際して制限時間を指定する

サイトやそのURLにアクセスするために、URLを元にhtmlを取得しています。サイトによっては反応時間が長かったり、サイトが存在せず待たされることがあります。そのようなときにhtmlを取得するまでの時間を制限することができます。

一般的なWebブラウザでは1分程度の時間を設けているようですが、秒単位で指定します。デフォルトは10秒です。

⑤アイテムをキーワード名サブフォルダ（親フォルダ）に一括保存する

アイテムの保存フォルダへの保存方法を指定します。ダウンロードされたアイテムは、親フォルダにベタで一括保存されます。

「Title名孫フォルダに分散保存」にチェックすると、URLペインのURLごとにhtmlのTitleをフォルダ名として子フォルダを形成しアイテムを保存します。第2Layer以降のURLに対してはTitleをフォルダ名として孫フォルダをツリー状に形成しアイテムを保存します。

Titleが得られないときは、NoTitle が代用されます。複数 NoTitle が発生する場合追い番で区別されます。

デフォルトは「チェック有り」です。

⑥Title名孫フォルダを階層しツリー状に形成する

第2Layer以降の保存方法を選択します。通常、URLの階層に準じてフォルダを作成してツリー状に孫、曾孫、玄孫フォルダを形成してアイテムを保存します。

「階層に関係なく平坦に保存」にチェックした場合、曾孫以下のフォルダは形成せずにすべて孫フォルダの階層に平坦にフォルダを形成しアイテムを保存します。

デフォルトは「チェック有り」です。

⑦htmlのTitleにKeywordの第一語句が含まれていたらダウンロード

Keywordを使ってURLを絞り込む機能です。Keywordの第一語句、スペースでKeywordを区切ったときの先頭の語句、がアクセス先のhtmlに記述されたTitleの中に含まれていたら、そのサイトのアイテムをダウンロードします。

この機能を適用する階層深度を指定します、階層の1から指定階層迄が適用範囲です。この機能を使わない場合は階層深度指定を **0** に設定します。

⑧ダウンロードしたURLのショートカットの扱い

アイテムを取得したサイトのURLショートカットを残すか否かの選択です。通常、URLのショートカットは残しません。

「対応フォルダにURLを保存」にチェックした場合、URLショートカットをアイテムを保存したフォルダに保存します。

デフォルトは「チェック有り」です。

⑨アイテムダウンロード制限時間の指定

アイテムをダウンロードするときの制限時間を分単位で設定します。アイテム毎には設定できず共通です。

デフォルトは3分です。

5.3 動作条件設定

Sniper条件設定

- ・KeyWordモード or バッチURLモードで探索 ☐ KeyWordFileモードで探索
- ・検索ページ数とLayer深度(URLペインが1) 2 ページ 2 階層
- ・検索1ページ当たりの取得URL数 ☐ 10 ☒ 20 ☐ 50
- ・同時にダウンロードするチャンネル数 ☐ 4 ☒ 8 ☐ 16
- ・Item無しURLが指定数になるまでサイト維持 ☒ 10 % 10 回で放棄
- ・既読URLが続いてもアクセスを継続 ☒ 10 回でサイトを放棄
- ・第2階層以降のURLの最短文字数指定 ☐ 30 文字以上に制限
- ・ダウンロードアイテムが少なくてもサイト継続 ☐ 5 未満連続で放棄
- ・Web Titleをフォルダ名に利用、最大文字数は 40 文字に制限する
- ・Web検索の間隔に制限を設けない ☐ 最少1分間隔に制限する
- ・ページ直貼りアイテム(<... src=)の扱い ☐ 直貼りもダウンロードする
- ・過去アクセス済みURLとの重複を許容する ☒ 過去分との重複を回避する
- ・先頭ItemURLだけ拡張子偽装チェック&解除 ☐ 全ItemURLをチェック&解除
- ・探索から外すサイトの指定 ☐ これらも同等に探索する
- ・URLペインのURLの階層下を探索範囲とする ☒ 同サイト上位階層も探索
- ・URLペインのURLのサイト内のみを探索する ☐ サイト外のURLも探索する
- ・Layer2以降末尾がhtml()のURLのみアクセス ☒ すべてのURLを探索する
- ・html取得に際してのサイトアクセス時間制限 10 秒でアクセス中止
- ・ItemをKeyWord名親フォルダに一括保存する ☒ title名フォルダに分散保存
- ・Title名フォルダを階層化しツリー状に形成する ☒ 階層に関係なく平坦に形成
- ・html TitleにKeyWord第一語句あればアクセス 2 階層迄適用(OでOff)
- ・アイテムをダウンロードしたURLの扱い ☒ 対応フォルダにURLを保存
- ・Itemダウンロード時間制限(時間超過で放棄) 1 分(全Item共通)

① フォントの設定 MS UI Gothic 9.75

② アイテムペインのURLをダブルクリックして画像を表示する時に用いるビューアー

③ 重複Table消去 ④ 設定して閉じる ⑤ 出荷状態に戻す ⑥ キャンセル

①フォントの設定

右のフォルダボタンをクリックして使用するフォントを設定します。フォルダ名、キーワード、拡張子、URL、ステータスバーのフォントを変更できます。

デフォルトは「MS UI Gothic」です。

②アイテムペインの画像URLをダブルクリックして表示するときに用いるアプリ

アイドリング中にアイテムペインとのURLが画像の場合ダブルクリックしてビューアーに表示することができます。そのときに用いるビューアーをフルパスで設定します。右のフォルダボタンをクリックするとファイルビューアーが開きますのでビューアーを指定します。

③「重複Table消去」ボタン

URL重複回避のためのテーブルを消去するボタンです。

「重複Table消去」をクリックしても設定画面を閉じるときにキャンセルをクリックするとTableは消去されません。消去するときは設定画面を閉じるときに、「設定して閉じる」ボタンをクリックします。

④「設定して閉じる」ボタン

この画面で設定した内容を反映して画面を閉じるするボタンです。

⑤「出荷状態に戻す」ボタン

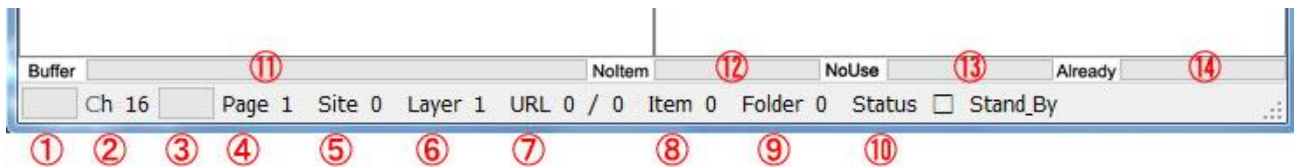
設定を出荷状態に戻すことができます。ただし、フォントの設定と外部ビューアーの設定はそのまま残り別途扱いを訪ねてきます。

⑥「キャンセル」ボタン

現在の設定内容を破棄して画面を閉じるボタンです。重複Tableの消去もキャンセルされます。

6. ステータスバー表示

画面最下部に動作状態を示すステータスバーが表示されます。



①アクションインジケータ

緑のバーが伸縮して動作していることを示します。それだけです、バーの長さなどに意味はありません。

②ダウンロードチャンネル数

設定されている最大ダウンロードチャンネル数を示します。

③稼働中のチャンネル数表示

稼働しているダウンロードチャンネルに応じてプログレスバーが表示されます。

④Web検索ページ番号

Web検索の何ページ目かを表します。

⑤アクセス中サイト番号

現在探索しているURL（反転表示されます）がURLペインのURLリストの何番目かを示します。

⑥アクセス中のLayer番号

URLペインのURLを第1階層（Layer=1）として、現在アクセスしているURLの階層位置を示します。

⑥アクセス中の階層のURL

現在アクセス中の階層にあるURL総数とアクセス中のURLの順番を示します。（現在の順番 of URL総数）

⑧ダウンロード済みアイテム数

指定したキーワードの探索でダウンロードしたアイテム総数を示します。

⑨作成したフォルダ数

指定したキーワードの探索で作成したフォルダの総数を示します。

⑩動作フェーズ

動作中のフェーズ略称を示します。

停止状態で"Stand_By"が表示され、動作中は

"Stand_By"、"Search_URLs"、"Access_URL
"、"Get_URLs"、"Get_Html""get_ItemURL"、"do_Download"、
"Already_URL"、"Avoid_URL"、"Buffer_Full"
、"No_Items"、"404Error"、"TimeOver"、"Obtained"、"Outer_URL"

などが表示されます。（瞬間表示で読み取ることができないものもあります。）

⑪ダウンロードバッファ残量計

条件に合致してダウンロードバッファに格納されたItemURLの数量を示します。違った言い方をすると、まだダウンロードされていないアイテムの残量を示します。

⑫アイテム無しURL連続件数

URLペインの登録URLにアクセスして得られるURLにアイテムが無かったとき、アイテム無しURLとしてカウントします。このプログレスバーはアイテム無しURLが続いている間カウントアップされます。

Max値は、設定の 5.1 ⑥「Item無しURLが指定数になるまでサイト維持」蘭で指定した継続許容回数となります。

Max値に達したとき、この設定蘭のチェックがOnであればそのサイトへのアクセスは中断され、次のURLへ移行します。

カウントアップしているときにアイテムが得られたらカウントはリセットされます。

⑬条件を外れたURL連続件数

5. 2の①、②、③の条件に外れるとアクセスされず「NoUse」としてカウントします。このプログレスバーは「NoUse」の進捗状況を表示します。

Max値は、設定の 5.1 ⑥「Item無しURLが指定数になるまでサイト維持」蘭で指定した継続許容回数の2倍の値となります。

Max値に達したとき条件設定されていればそのサイトへのアクセスは中断され、URLペインで次のURLに移行します。

⑭既読URL連続件数

アイテム無しURL連続件数と同じように、既読URLをカウントします。Max値は既読の継続許容回数となり、設定の 5.1 ⑦「既読URLが続いてもサイトへのアクセスを継続」蘭で指定した継続許容回数となります。

Max値に達したとき、この設定蘭のチェックがOnであればそのサイトへのアクセスは中断され、URLペインで次のURLへ移行します。

カウントアップしているときに、新しいURLに遭遇するとカウントはリセットされます。