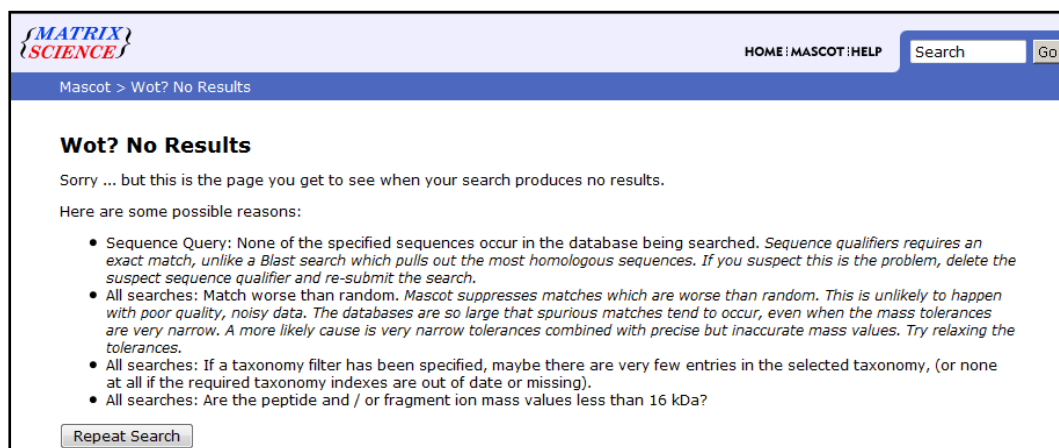


# speclist.txt ファイルフォーマット変更に伴う SwissProt データベースの再構築と設定変更について

## 1. はじめに

MASCOT にて Taxonomy 設定に利用していた **speclist.txt** ファイルのフォーマットが変更されたため、SwissProt のバージョン 2011\_06 更新後、**taxonomy** を選択して **SwissProt** で検索をしても結果がでない (**Wot? No Results** と表示される、**下図**) 問題が発生しています。



今後リリース予定の MASCOT ver 2.4 以降ではこの問題に MASCOT プログラム側で対処する予定ですが、現在 MASCOT ver 2.3 以下をご利用の方が問題を解決するためには、**弊社で修正した speclist.txt を MASCOT サーバーの所定の場所にコピーし、SwissProt データベースを再度構築する必要があります**。また、問題解決後再び同じ問題が起きないように、speclist.txt の入手先を弊社サイトに変更するよう、設定を変更する必要があります。

本資料では修正版 speclist.txt ファイルの入手並びに SwissProt データベースの再構築方法（項目2）と、データベース自動更新プログラム db\_update.pl が speclist.txt を入手する先を EBI から弊社サイトへと変更する設定方法（項目3）について記しています。

最新の情報が弊社 WEB サイトにおいても公開されております。そちらも併せてご参照ください。

[http://www.matrixscience.com/help/seq\\_db\\_setup\\_Sprot.html](http://www.matrixscience.com/help/seq_db_setup_Sprot.html)

## 2. 修正版 speclist.txt ファイルの入手、並びに SwissProt データベースの再構築方法

ファイルの入れ替えとデータベースの再構築は、以下の手順で行います。

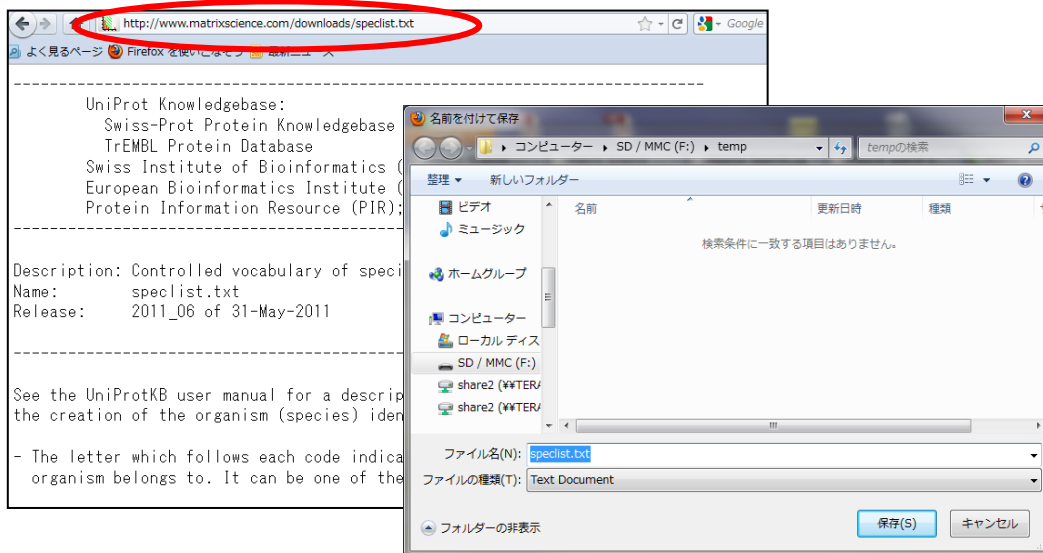
- |                                  |                         |
|----------------------------------|-------------------------|
| ① 修正版 <b>speclist.txt</b> ファイル取得 | ④ <b>.stats</b> ファイルの削除 |
| ② <b>speclist.txt</b> ファイルの入れ替え  | ⑤ <b>MASCOT</b> サービスの開始 |
| ③ <b>MASCOT</b> サービスの停止          | ⑥ データベース構築の確認、検索テスト     |

### ① 修正版 speclist.txt ファイル取得

以下の URL (弊社 WEB サイト) から **speclist.txt** ファイルを取得します。

<http://www.matrixscience.com/downloads/speclist.txt>

ウェブブラウザで直接上記 URL を開いた場合、テキストファイルが開かれた状態になると思いますが、「保存」「名前をつけてページを保存」などで当該ファイルをそのまま保存してください(下図)。



ブラウザで直接入力するのが苦手な方は、以下のリンクをたどりファイルを入手してください。

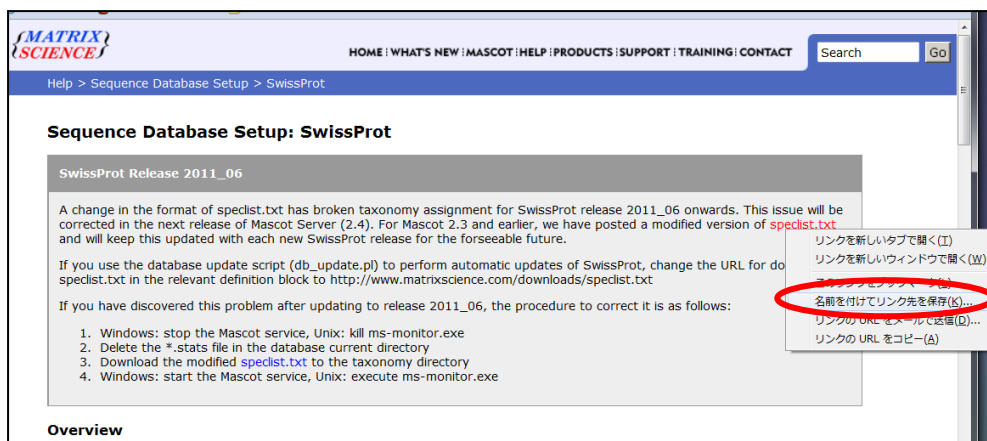
[www.matrixscience.com](http://www.matrixscience.com) のページ (キーワード 'mascot' 検索するとだいたいトップで出てきます)

→HELP→Sequence Database Setup → SwissProt

**\*ローカルの MASCOT Server でなく、公開 WEB サイトの方です。間違えないようにしてください。**

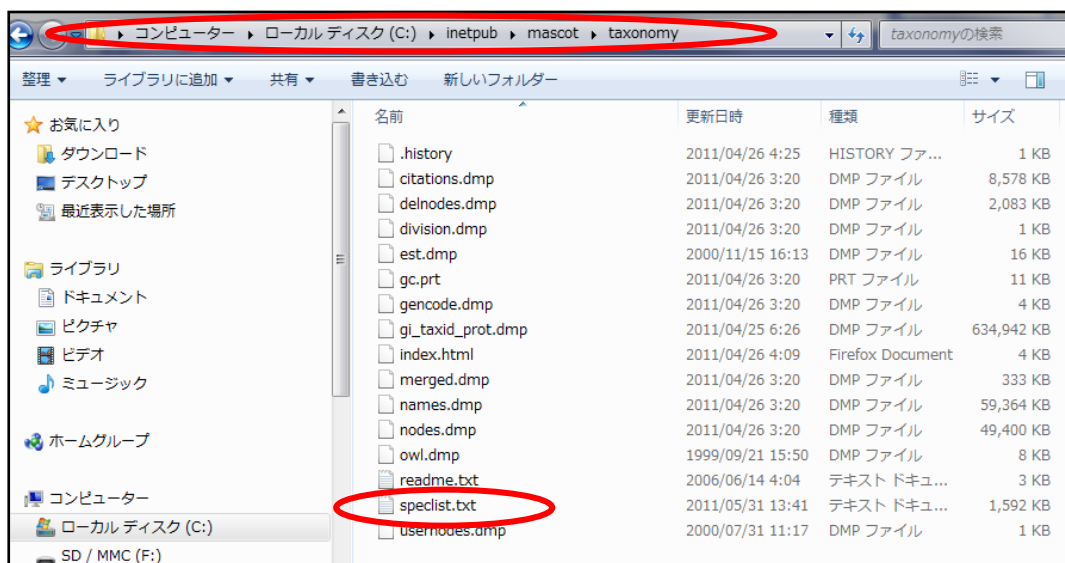
## speclist.txt ファイルフォーマット変更に伴う SwissProt データベースの再構築と設定変更について

画面上部の「speclist.txt」ハイパーリンクを右クリックし、「対象をファイルに保存」「名前をつけてリンク先を保存」を選びます（下図）。



## ② speclist.txt ファイルの入れ替え

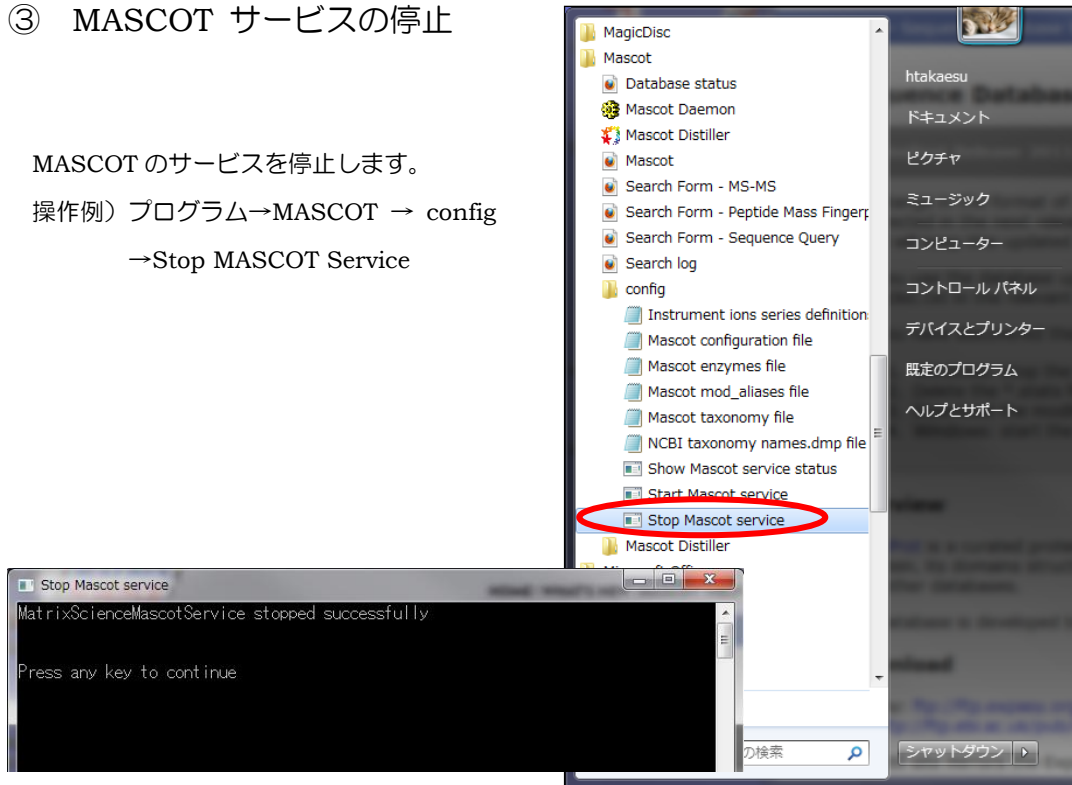
MASCOT サーバー上の、MASCOT がインストールされているフォルダ（通常 C:\netpub\mascot フォルダ）以下、taxonomy フォルダ内の speclist.txt ファイルについて、バックアップを取った上で、①で取得したファイルと入れ替えます（下図）。入れ替えは上書き置換などで行います。



### ③ MASCOT サービスの停止

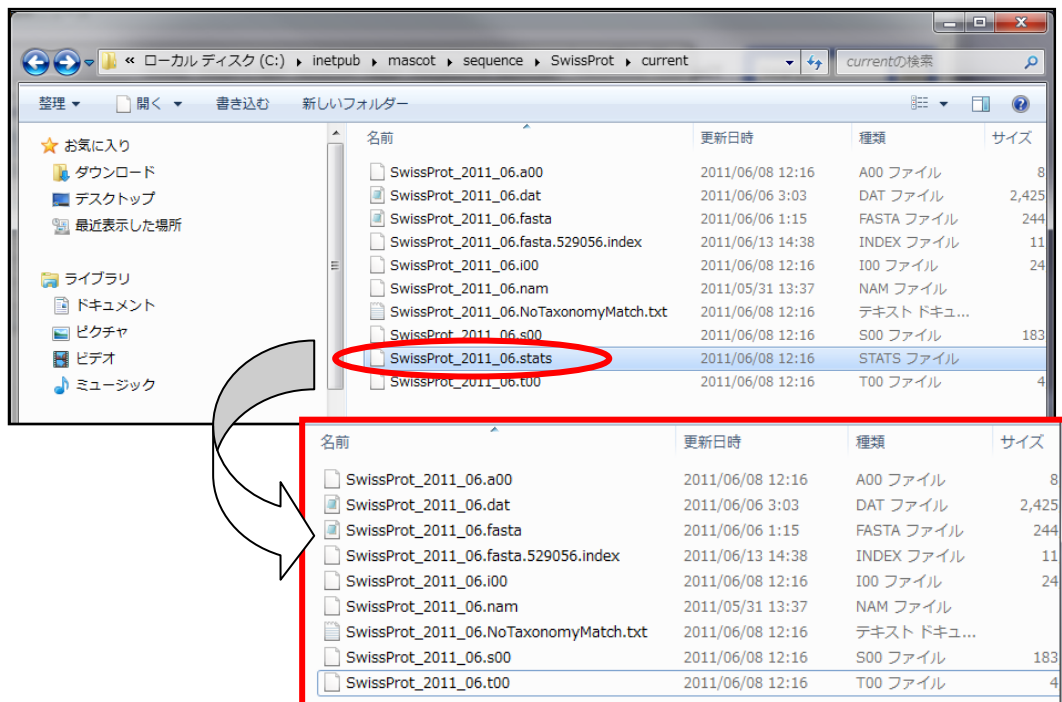
MASCOT のサービスを停止します。

操作例) プログラム→MASCOT → config  
→Stop MASCOT Service



### ④ .stats ファイルの削除

MASCOT サーバー上の、MASCOT がインストールされているフォルダ（通常 C:\netpub¥mascot フォルダ）以下、sequence¥SwissProt¥current フォルダを開き、**SwissProt\_2011\_06.stats** ファイルを削除します。

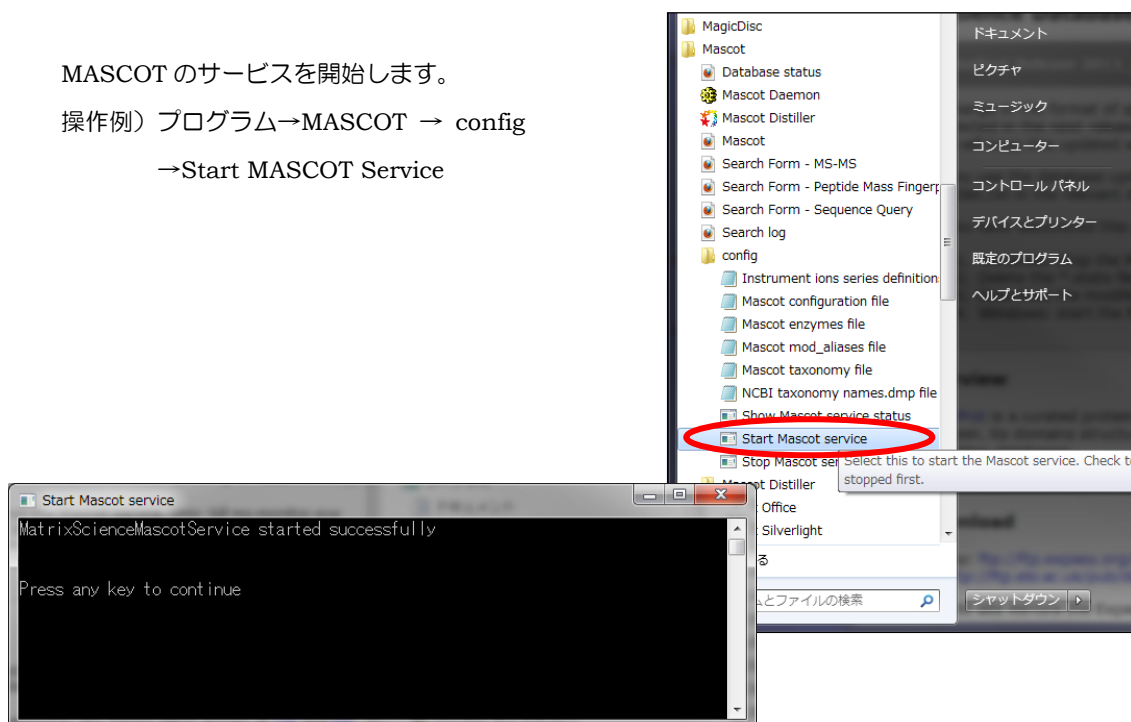


## ⑤ MASCOT サービスの開始

MASCOT のサービスを開始します。

操作例) プログラム→MASCOT → config

→Start MASCOT Service

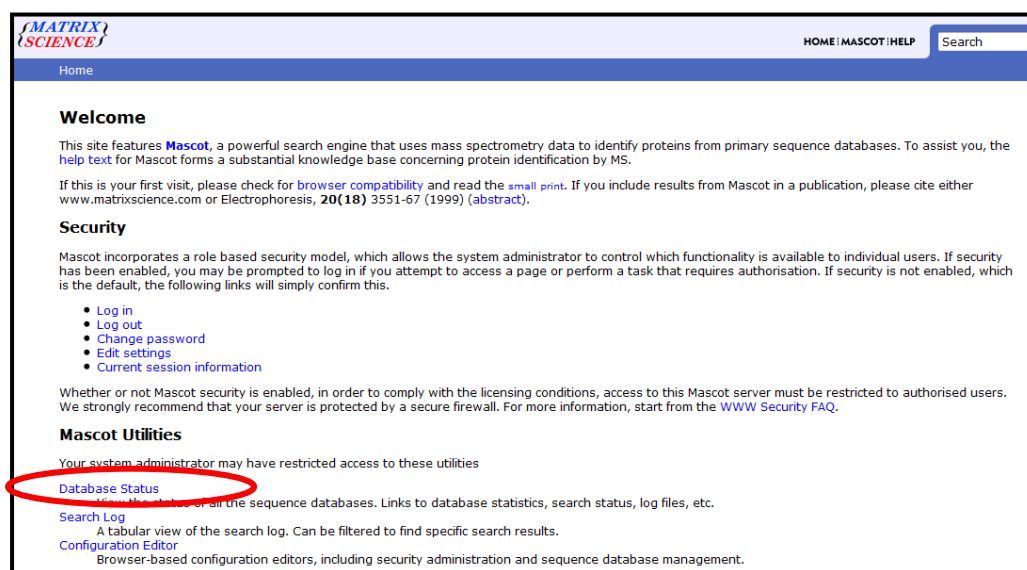


## ⑥ データベース構築の確認、検索テスト

サービス開始後、SwissProt データベースの再構築が始まります。再構築の状況を確認するためには、Database Status 画面を開き、SwissProt の Status 項目をご覧ください。

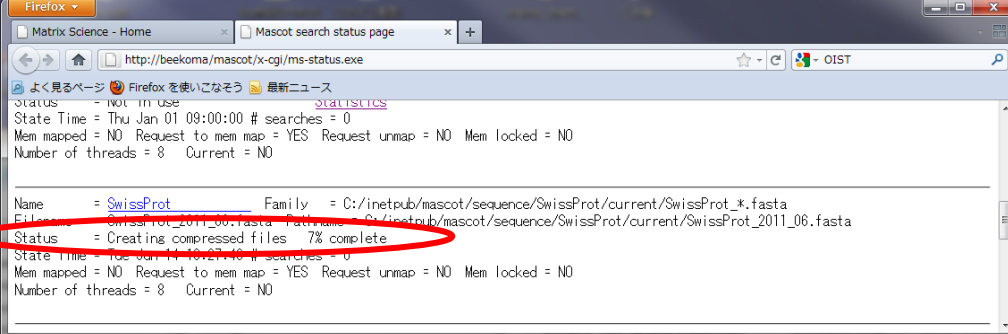
Database status 画面へのアクセス方法は以下の通りです。

MASCOT の Home 画面を開き、「**Database Status**」をクリックします。



## speclist.txt ファイルフォーマット変更に伴う SwissProt データベースの再構築と設定変更について

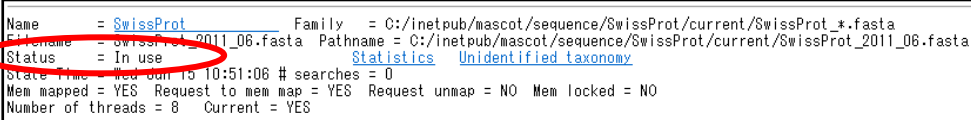
Name 項目が “SwissProt”の箇所を探します。SwissProt の中身に関する内容が書かれています。  
 Status 項目はデータベースの現状を表しており、サービス開始後まもなくは”Creating compressed files N % complete”となっているはずで、これはデータベース構築中であることを表しています。



```

Name = SwissProt Family = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta
File = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_2011_06.fasta
Status = Creating compressed files 7% complete
State Time = Thu Jan 01 09:00:00 # searches = 0
Mem mapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = 8 Current = NO
  
```

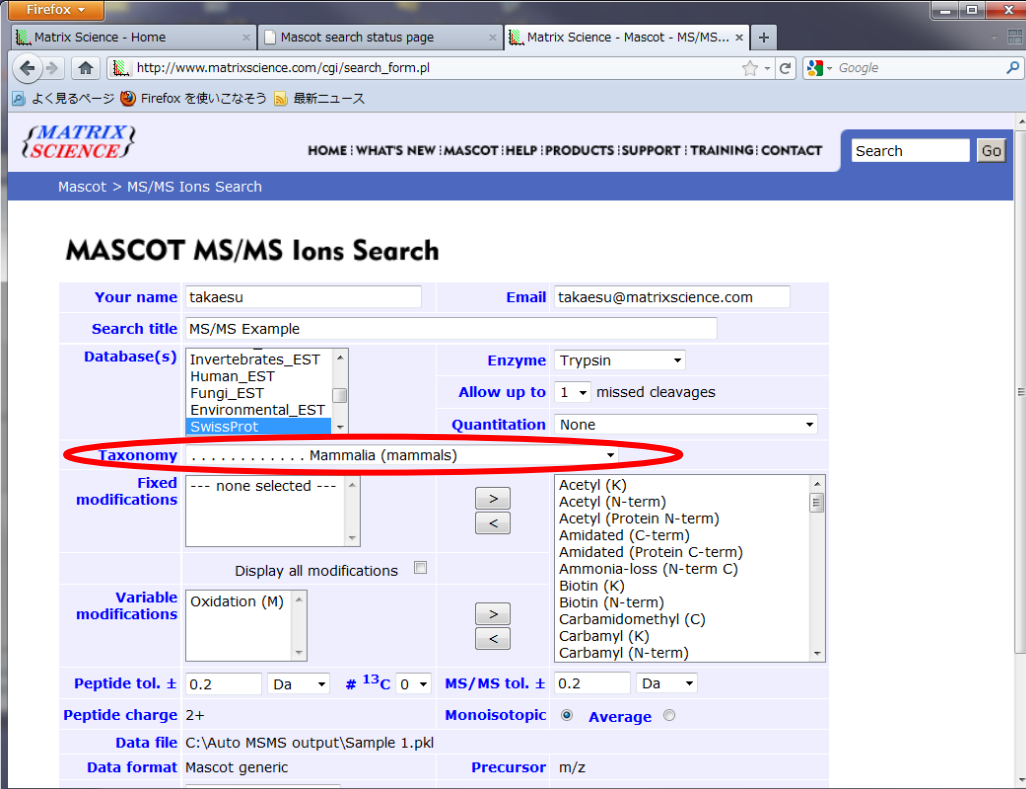
最終的に **Status 項目が”In Use”**となれば構築完了,使用可能です。



```

Name = SwissProt Family = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta
File = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_2011_06.fasta
Status = In use
State Time = Wed Jun 15 10:51:06 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = 8 Current = YES
  
```

Status が In Use になったのち、確認のため検索テストを行ってください。テスト時は何かしらの taxonomy を選択した上で検索を行ってください。



MASCOT MS/MS Ions Search

Your name: takaesu Email: takaesu@matrixscience.com

Search title: MS/MS Example

Database(s): Invertebrates\_EST, Human\_EST, Fungi\_EST, Environmental\_EST, SwissProt

Enzyme: Trypsin

Allow up to: 1 missed cleavages

Quantitation: None

**Taxonomy: Mammalia (mammals)**

Fixed modifications: none selected

Variable modifications: Oxidation (M)

Peptide tol. ±: 0.2 Da # 13C: 0 MS/MS tol. ±: 0.2 Da

Peptide charge: 2+ Monoisotopic: Average

Data file: C:\Auto MSMS output\Sample 1.pkl

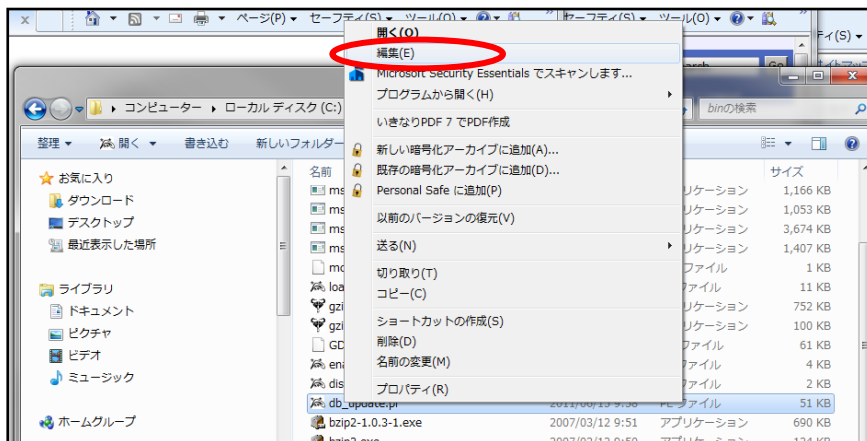
Data format: Mascot generic Precursor: m/z

### 3. データベース自動更新プログラム db\_update.pl が今後 EBI で公開している speclist.txt を取得しないようにする設定方法

項目2の処置を行いデータベースが正常に動作した後も、データベースの更新を行い **speclist.txt** が新しくなると再び同じ問題が起きてしまいます。今後問題を発生させないためにはデータベースの更新を止めるか、以下のようにデータベースの自動更新プログラムの内容を一部書き換えてください。

MASCOT サーバー上の、MASCOT がインストールされているフォルダ（通常 C:\netpub\mascot フォルダ）の **bin** フォルダを開き、**db\_update.pl** のバックアップを取った上で、db\_update.pl をテキストエディタで開きます。db\_update.pl をテキストエディタで開く方法は、ファイルを選択し右クリックして「編集」を選択するか、プログラムから開く→「メモ帳」または「ワードパッド」を選びます。

\* NotePad,WordPad どちらでも OK です。

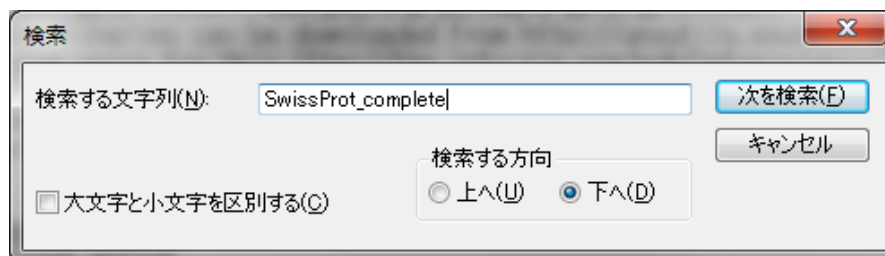


```
db_update.pl - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
#!/usr/local/bin/perl
#####
# Downloads sequence database and taxonomy file updates
# Sequence database is specified by keyword command line argument
# Requires wget (http://www.gnu.org/software/wget/wget.html)
# tar (http://www.gnu.org/software/tar/tar.html)
# gzip (http://www.gnu.org/software/gzip/gzip.html)
# Windows binaries can be downloaded from http://unutils.sourceforge.net/
# Info-ZIP unzip for Unix (ftp://ftp.info-zip.org/pub/infozip/unix/)
# unzip for Windows (ftp://ftp.info-zip.org/pub/infozip/win32/)
#
# Optional processing:
# varsplic to expand variants described in Swiss-prot annotations
# ftp://ftp.ebi.ac.uk/pub/software/swissprot/varsplic/
# ftp://ftp.ebi.ac.uk/pub/software/swissprot/Swissknife/
#####
# COPYRIGHT NOTICE
# Copyright 1998-2009 Matrix Science Limited All Rights Reserved.
#
# This program may be used and modified within the licensee's organisation
# provided that this copyright notice remains intact. Distribution of this
# program or parts thereof outside the licensee's organisation without the
# prior written consent of Matrix Science Limited is expressly forbidden.
#####
# $Archive:: /www/bin/db_update.pl $ #
# $Author: johnc $ #
# $Date: 2010-05-02 17:49:58 $ #
# $Revision: 1.34 $ #
# $NoKeywords: $ #
```

## speclist.txt ファイルフォーマット変更に伴う SwissProt データベースの再構築と設定変更について

SwissProt 更新時に speclist.txt ファイルを取得するサイトを、EBI から MatrixScience 社のサイトに  
変更します。

ファイル取得の記述箇所を見つけるため検索機能を使います。ファイルを開いた後、Ctrl+F またはメ  
ニューバーから「検索」を選びます。検索欄に、「**SwissProt\_complete**」と入力し、検索を実行しま  
す。



検索後最初に SwissProt の更新に関する記述場所が見つかります。

```
db_update.pl - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
# SwissProt_complete_from_EBI
#
} elsif ($ARGV[0] eq "SwissProt_complete_from_EBI") {
  $db_name = "SwissProt";
  $local_incoming_directory = "$SEQUENCE/SwissProt/incoming";
  $local_current_directory = "$SEQUENCE/SwissProt/current";
  $fasta_file_url =
"ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz";
  $name_file_url = "ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete/reldate.txt";
  $reference_file_url =
"ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.dat.gz";
  $taxonomy_file_url[0] = "ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz";
  $taxonomy_file_url[1] =
"ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete/docs/speclist.txt";
  $version_regex = "Swiss-Prot%s+.*Release%s+([\d_]+)";
#
# SwissProt_varsplc_from_EBI
#
} elsif ($ARGV[0] eq "SwissProt_varsplc_from_EBI") {
  $db_name = "SwissProt";
```

\$taxonomy\_file\_url[1] をご覧ください。取得するファイルが

**"ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current\_release/knowledgebase/  
complete/docs/speclist.txt";**

となっていますが、これを

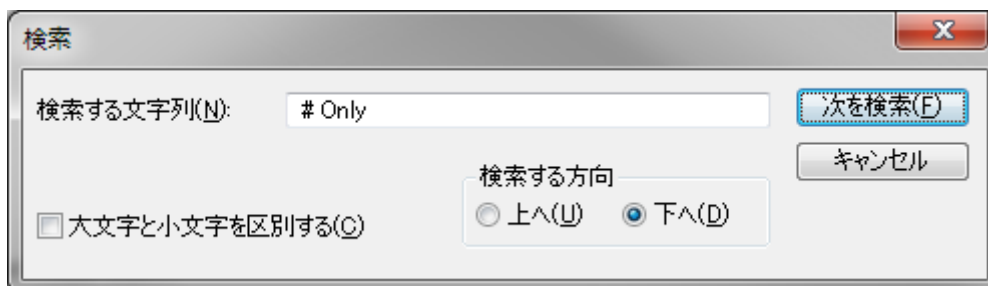
**"http://www.matrixscience.com/downloads/speclist.txt";**

に変更します。特に、**行最後の';'**を忘れないよう、ご注意ください。

```
db_update.pl - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
# SwissProt_complete_from_EBI
#
} elsif ($ARGV[0] eq "SwissProt_complete_from_EBI") {
  $db_name = "SwissProt";
  $local_incoming_directory = "$SEQUENCE/SwissProt/incoming";
  $local_current_directory = "$SEQUENCE/SwissProt/current";
  $fasta_file_url =
"ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz";
  $name_file_url = "ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete/reldate.txt";
  $reference_file_url =
"ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.dat.gz";
  $taxonomy_file_url[0] = "ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz";
  $taxonomy_file_url[1] = http://www.matrixscience.com/downloads/speclist.txt;
  $version_regex = "Swiss-Prot%s+.*Release%s+([\d_]+)";
#
# SwissProt_varsplc_from_EBI
#
} elsif ($ARGV[0] eq "SwissProt_varsplc_from_EBI") {
  $db_name = "SwissProt";
  $local_incoming_directory = "$SEQUENCE/SwissProt/incoming";
```



さらに続けて、「# Only」を検索します。「#」と「Only」の間にはスペースが入りますのでご注意ください。



検索後表示される行「 # Only use --ignore-length option if desperate」の1つ上の行、「 \$sysCall .= " --directory-prefix=\$download\_directory";」を書き換えます。元々あった行の先頭に「#」をつけてコメントアウトした後、以下の例のような6行を追加します。行は直接入力か、コピー&ペーストで貼り付けてください。難しく感じられるようでしたら、弊社の技術サポートまでご連絡ください。ご連絡頂く際には、MASCOT の標準フォルダ(デフォルトは C:\inetpub\mascot ) の場所も合わせてご連絡ください。

#### 【変更前】

```
$sysCall .= " --directory-prefix=$download_directory";
# Only use --ignore-length option if desperate
# if ($url =~ /^http:/i) {
```

#### 【変更後】

```
# $sysCall .= " --directory-prefix=$download_directory";

if ($remoteName=="speclist.txt") {
    $sysCall .= " -O $download_directory"."/$remoteName";
    $sysCall =~ s#¥¥##/g;
}else {
    $sysCall .= " --directory-prefix=$download_directory";
}

# Only use --ignore-length option if desperate
# if ($url =~ /^http:/i) {
```

最終的には次ページ図のようになるかと思います。

(図は挿入場所を分かりやすくするため###などの行も追加されています。)

## speclist.txt ファイルフォーマット変更に伴う SwissProt データベースの再構築と設定変更について

```

db_update.pl - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
}
# download the file
$sysCall = wget;
if ($ENV{'WINDIR'}){
  $sysCall = "s#/#%#%#g;
}
while (($key, $value) = each(%wget_options)) {
  if (length($value)) {
    $sysCall .= " $key=$value";
  } else {
    $sysCall .= " $key";
  }
}
# seems like wget cannot cope with --retr-symlinks and --timestamping
if ($matchLineSave =~ /%s+($matchName)%s*$/) {
  $sysCall .= " --timestamping";
} else {
  $sysCall .= " --retr-symlinks";
}
####edited by takaesu
# $sysCall .= " --directory-prefix=$download_directory";

if ($remoteName=="speclist.txt") {
  $sysCall .= " -O $download_directory ./$remoteName";
  $sysCall = "s#/#%#%#g;
} else {
  $sysCall .= " --directory-prefix=$download_directory";
}
####

```

変更後、ファイルを上書き保存します。

```

db_update.pl - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
新規(N) Ctrl+N
開く(O) Ctrl+O
上書き保存(S) Ctrl+S
名前を付けて保存(A)...
ページ設定(U)...
印刷(P)... Ctrl+P
メモ帳の終了(X)
$taxonomy_file_url[0] = "es/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz";
$taxonomy_file_url[1] = "ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete/rel date.txt";
$version_regex = "Swiss-Prot%+.*Release%+([%d.]+)";
#
## SwissProt_varsplc_from_EBI
#
] elsif ($ARGV[0] eq "SwissProt_varsplc_from_EBI") {
$db_name = "SwissProt";
$local_incoming_directory = "$SEQUENCE/SwissProt/incoming";

```

作業はこれで終了です。更新プログラムがうまく動くかどうかについては、1カ月後のデータベース更新にてご確認ください。

資料は以上です。

ご不明点などありましたら技術サポート担当者にご連絡ください。

マトリックスサイエンス株式会社

〒101-0021 東京都千代田区外神田 6-10-12 KNビル3階

電話：03-5807-7897

電子メール：support-jp@matrixscience.com