

配列データベースのセットアップ

1 概要

MASCOT Server は FASTA フォーマットで記述されたアミノ酸配列または塩基配列データベースに対して MASCOT 検索を実行します。配列データベースに関する設定は、MASCOT Server の設定ファイルである `mascot.dat` に定義されています。配列データベースのセットアップ状況、統計情報、稼働状況等は、次の URL (MASCOT search status page 画面) で確認することができます。

<http://MASCOTServerホスト名/mascot/x-cgi/ms-status.exe>

MASCOT Server システムに十分な量のメモリが搭載されている場合は、配列データベース・ファイルをメモリ上にマップすることにより、より高速に MASCOT 検索を実行することができます。

標準でインストールされる `db_update.pl` プログラムを利用することにより、配列データベース・ファイルのダウンロードから更新までを自動化することができます。また、MS Windows OS のタスクスケジューラや Linux/Unix の Cron プログラムを利用することにより、指定した日時に (周期的に) `db_update.pl` プログラムを実行することができます。

新規にデータベースをセットアップする場合は、配列データベースを構成する各エントリからアクセッション番号などの文字列を抽出するための「正規表現」の知識が必要になります。

配列データベースは 64 個までセットアップすることができます。

1-1 FASTA フォーマット

FASTA フォーマットは、1 行目はケット (大なり記号、>) で始まるタイトル行、2 行目以降は 1 文字コード表記によるアミノ酸または塩基の配列行で構成されたエントリをひとつの単位として、このエントリが複数個集合した形式を持っています。たとえば、MSDB 蛋白質データベースの場合は、2004 年 9 月現在で 150 万件以上のエントリで構成されています。配列データベース・ファイルはアミノ酸配列または塩基配列で構成することができますが、両者が混在した構成にすることはできません。塩基配列データベース・ファイルの場合は、3 個の塩基をリアルタイムにアミノ酸に変換しながら 6 種類の読み枠 (six reading frames) に対して MASCOT 検索を実行します。

FASTA フォーマットのタイトル行はケット (>) で始まり、1 つ以上のアクセッション番号 (Accession number)、説明 (Definition/Description)、生物分類 (Taxonomy) などの文字列が続きますが、その文法・書式は、行頭にケット (>) を配置すること以外は配列データベースによってまちまちです。タイトル行と配列行は改行コードで区切られます。

配列行に含まれる配列文字数は配列データベースによって異なりますが、通常 60～1000 文字以上です。MASCOT Server が認識できる配列行中の配列文字数は、デフォルト設定では 50000 です。次の行がタイトル行かまたはファイルの終わりで配列行の最後を認識しています。

FASTA フォーマットで記述されたエントリの例 (MSDB と EST_human) を示します。

FASTA フォーマットで記述されたエントリの例

1. MSDB データベース (アミノ酸配列) の例

```
>A27187 ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome c1 precursor - Neurospora crassa
MLARTCLRSTRTFASAKNGAFKFAKRSASTQSSGAAESPLRLNIAAAAATAVAAGSI AWYYHLYGFASAMTPAEGLHATKYPWVHEQWLKTF
DHQALRRGFQVYREVCASCHSLSRVPYRALVGTILTVDEAKALAEENEYDTEPNQGEIEKRPGLSDYLPDPYKNDEAARFANNGALPPDLSL
IVKARHGGCDYIFSLLTGYPDEPPAGASVAGLNFNPYFPGTGIAMARVLYDGLVDYEDGTPASTSQMAKDVVEFLNWAEEPMDRKRMGMKV
LVVTSVLFALSYYVKRYKWAWLKSRIIVYDPPKSPPPATNLALPQQRKS
```

2. EST_human データベース (塩基配列) の例

```
>gi|272212|gb|M61961.1| EST00010 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone
HFBA07, mRNA sequence
ATTAATAGGAAAGATGATTGTATAGATGGTGGGCTATTAAGTACAGATCAGGATGAGAATCGGGAGTGCCTTTACATGTGTGGTACCCAAATGGG
TGTTGGATATAAGAGTAACAAAAGGACTGAAAGGGTTAAAAAGAAAAGAAAAAAACTCCCTGGTTGGGAGGGTGTAAAGTATCGAGTG
TTTTCCAAACCATTCCTCCTCTGCTCACCTACCCCTAGGTATTAAGGAGATAACTTTTAAAAAAGAAAGAAATTGGCTCAAAGGTACTGTAA
ATTCTAGGATTATATACCTTTATATAGGTTTCATTCCT
```

MSDB データベースでは、ケット (>) に続く A27187 がアクセス番号、ubiquinol から precursor までが説明、Neurospora crassa が生物分類になります。また、EST_human データベースでは、gi|272212 と gb|M61961.1 の2つのアクセス番号を持ち、EST00010 以降が説明になります。gi|272212 は NCBI が管理しているアクセス番号ですが、このエントリは GenBank データベースの gb|M61961.1 に対応していることを示しています。

FASTA フォーマットの配列データベース・ファイルは上の例に示したエントリが複数個集合した形式になっています。Swiss-Prot データベース (部分) の FASTA ファイルの例を示します。

FASTA フォーマットで記述された配列データベース (Swiss-Prot) の例

(省略)

```
>10KD_VIGUN (P18646) 10 KD PROTEIN PRECURSOR (CLONE PSAS10)
MEKKSIAGLCFLFLVFAQEVVVQSEAKTCENLVDTYRGPCTTGGSCDDHCKNKEHLLSGRCRDDVRCWCTRNC
>110K_PLAKN (P13813) 110 KD ANTIGEN (PK110) (FRAGMENT)
FNSNMLRGSVCEEDVSLMTSIDNMIIEIDFYEKEIKYKSHSGGVIKGMDYDLEDDENDEMEQMVVEEVADHI TQDMI DEVAHHVLDNI THDM
```

```
AHMEEIVHGLSGDVTQIKEIVQKVNVAVEKVKHIVETEETQKTVEPEQIEETQNTVEPEQTEETQKTVEPEQTEETQNTVEPEQIEETQKTVEP
EQTEEAQKTVEPEQTEETQKTVEPEQTEETQKTVEPEQTEETQKTVEPEQTEETQKTVEPEQTEETQKTVEPEQTEETQKTVEPEQTEETQNTV
EPEPTQETQNTVEP
```

```
>11S3_HELAN (P19084) 11S GLOBULIN SEED STORAGE PROTEIN G3 PR
```

```
MASKATLLLAFTLLFATCIARHQQRQQQNCQLQNI EALEPIEVIQAEAGVTEIWDAYDQQFQCAWSILFDTGFNLVAFSCLPTSTPLFWPSS
REGVILPGCRRTYEYSQEQQFSGEGRRGGEGTFRTVIRKLENLKEGDVVAIPTGTAHWLHNDGNTELVVVFLDTQNHENQLDENQRRFFLAG
NPQAQAQSQQQQRQPRQQSPQRQRQRQGGQGNAGNIFNGFTPEL I AQSFNVDQETAQKLQGGNDQRGHI VNVGQDLQ I VRPPQDRRSPRQQ
EQQATSPRQQEQQQRRGGWSNGVEET I CSMKFKVNI DNPSQADFVNPQAGSI ANLNSFKFPI LEHLRLSVERGELRPNA I QSPHWT I NAHNL
LYVTEGALRVQI VDNQGNVSFDNELREGQVVV I PQNFAV I KRANEQGSRWVSFKTNDNAM I ANLAGRVSASAASPLTLWANRYQLSREEAQLK
FSQRETVLFFAPSF SRGG I RASR
```

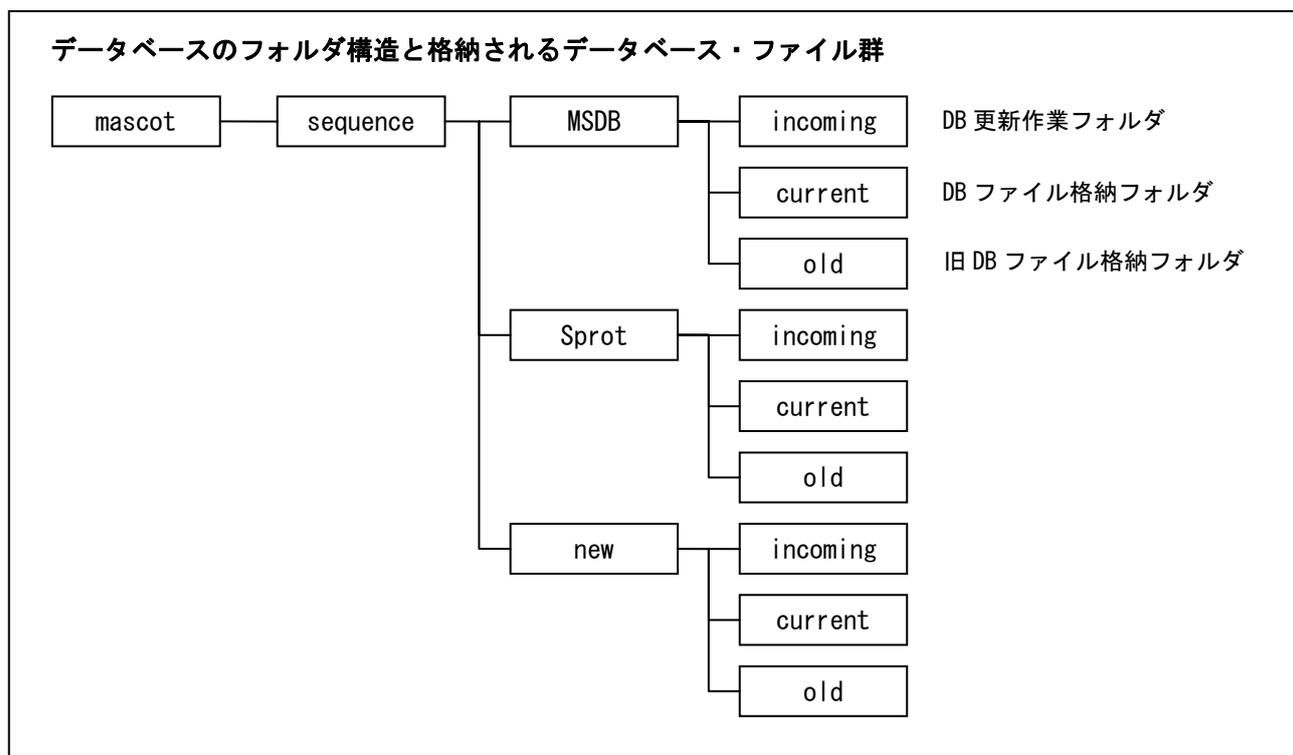
(省略)

FASTA フォーマットで記述されていない配列データベースの場合は、FASTA フォーマットに変換する必要があります(通常は簡単な変換プログラム作成し、変換します)。

Mascot Server は FASTA フォーマットの配列データベース・ファイルに対し、直接 MASCOT 検索を実行するわけではありません。Mascot Monitor プログラム (ms-monitor.exe) が配列データベース・ファイルを構成する各エントリをアクセス番号、説明、生物分類、配列に分割・圧縮することにより、メモリを節約すると同時に検索速度の向上を図っています。短い配列で構成された配列データベースや塩基配列データベースの場合は、より効果的にメモリを節約することができます。

1-2 フォルダ構造

一般的に配列データベースは定期・不定期に更新されることを想定して、MASCOT Server が使用する配列データベース・ファイル群は次の構造を持つフォルダに格納され、管理されます。各フォルダの役割及び格納される配列データベース・ファイル群は次の通りです。なお、フォルダ名や配列データベース・ファイル名を構成する文字列は大文字、小文字が区別されます。また、スペース文字は使用できませんので注意が必要です。



(1) mascot フォルダ

MASCOT Server のインストールフォルダです。インストール時に C:\inetpub フォルダの下に作成されます。

(2) sequence フォルダ

MASCOT Server が一意に認識できる配列データベース名を持つフォルダを格納します。

(3) MSDB、Sprot、new フォルダ

MASCOT Server が一意に認識できる配列データベース名を持つフォルダです。このフォルダの中に配列データベースに関連するファイル群が格納される3つのフォルダ (current、incoming、old) を作成します。

(4) current フォルダ

MASCOT 検索で使用する配列データベースの実体である FASTA フォーマットの配列データベース・ファイルや Mascot Monitor プログラムが作成した圧縮ファイル群が格納されます。

(5) incoming フォルダ

配列データベースを更新する際に、ダウンロードされる圧縮された配列データベース・ファイルの保存や解凍などの作業領域として使われます。

(6) old フォルダ

更新された配列データベースがセットアップされ有効になった後に不要になった古い配列データベース・ファイル (FASTA ファイル) が格納されます。すでに存在するより古い配列データベース・ファイルは削除されます。

1-3 データベース・ステータス

MASCOT Serverにセットアップされた配列データベースはMASCOT検索に対して有効/無効になるように設定することができます。MASCOT検索に対して有効になっている配列データベースのステータス (稼働状況) は、<http://MASCOTホスト名/mascot/>で表示される画面の中の、[Database status](#)リンクをクリックするか、

<http://MASCOTホスト名/mascot/x-cgi/ms-status.exe>

画面 (MASCOT search status page) で確認することができます。また、画面中の配列データベース名のリンクをクリックすると、その配列データベースに対して現在実行されているジョブの一覧を確認することができます。

The screenshot displays two overlapping browser windows. The top window, titled 'Mascot search status page - Microsoft Internet Explorer', shows the main status page with the following text:

```
Version: 2.0.05 - Licensed to: Matrix Science in-house demonstration system (Tokyo), (4 processors).
1 processor available, (number 0). Using processor number 0. [0 searches running]
Search log monitor log error log Error message descriptions Do not auto refresh this page
```

Below this, it lists database information for MSDB:

```
Name = MSDB Family = C:/INETPUB/MASCOT/sequence/MSDB/current/MSDB_*.fasta
Filename = MSDB_20040630.fasta Pathname = C:/INETPUB/MASCOT/sequence/MSDB/current/MSDB_20040630.fasta
Status = In use Statistics Unidentified taxonomy
State Time = Fri Oct 15 16:08:53 # searches = 0
Mem mapped = NO Request to mem map = NO Request unmap = NO Mem locked = NO
Number of threads = 4 Current = YES
```

The bottom window, titled 'Mascot database status - Sprot', shows a table of jobs:

Current jobs					
Job	PID	Start_Time	Dur.	Status	User
1418	352	Fri Oct 15 16:27:45	6	Searching...	test

Completed jobs					
Job	PID	Start_Time	Dur.	Status	User
1412	1480	Fri Oct 15 16:26:49	6	User read res	test
822	1468	Fri Oct 15 16:27:08		User read res	GETTAX
821	1488	Fri Oct 15 16:27:08		User read res	GETSEQ
820	1468	Fri Oct 15 16:27:08		User read res	GETSEQ
819	1472	Fri Oct 15 16:26:29		User read res	GETSEQ

At the bottom of the second window, there is a link: [Purge unfinished jobs in list for this DB](#) and another link: [Back to main status page](#).

1-4 設定ファイル mascot.dat

MASCOT Server は様々な設定値を持っていますが、これらの設定値は mascot.dat ファイルにテキストで記載されていますので、MS Windows OS に付属しているワードパットのようなテキストエディタで開き、編集することができます。

mascot.dat ファイルは 8 個セクションから構成されており、各セクションはセクション名キーワードで始まり end キーワード終わります。セクション名キーワードは、1. Databases、2. PARSE、3. WWW、4. Taxonomy_n (n=1、2、・・・)、5. Cluster、6. UniGene、7. Options、8. Cron です。データベースの設定に関する主なセクションは、次の 4 つです。

行頭にコメント (#) が存在する場合、その行はコメント行として無視されます。

(1) Databases

登録する配列データベース各々に関し、その属性を 1 行で定義します。配列データベースの属性は、配列データベースの名称、配列データベース・ファイルの完全パス (フォルダを含めたファイル名)、アミノ酸配列または塩基配列の指定、アクセス番号部分や説明部分に対応する文字列の抽出条件 (PARSE セクションで定義された抽出ルール番号)、メモリへのマップやロック設定フラグ、生物情報抽出条件 (Taxonomy_n で定義された抽出ルール番号) などで構成され、属性間はスペースかタブで区切ります。

行頭にコメント (#) を付けることにより、その配列データベースを MASCOT 検索に対して無効にすることができます。

(2) PARSE

配列データベース・ファイルの各エントリのタイトル行からアクセス番号や説明部分に対応する文字列を抽出するためのルールをルール番号とともに定義します。抽出ルールは正規表現 (Basic regular expression) を利用します。

(3) WWW

配列データベースのアクセス番号に対応する配列や詳細情報を抜き出すためのプログラムである ms-getseq.exe を利用するための設定です。詳細情報が Web サイトに存在する場合は、直接その Web サイトを指定することもできます。

(4) Taxonomy_n (n=1, 2, ...)

配列データベース・ファイルの各エントリが持つ生物分類を抽出するための条件を定義します。抽出条件は配列データベースの FASTA 形式に依存し、設定内容は若干複雑です。

2 配列データベースの更新

NCBI などが管理している配列データベース・ファイルはエントリの追加や変更が施され、新しいリリースとして定期的に更新されます。MASCOT Server にセットアップされ、MASCOT 検索に対して有効になっている配列データベース・ファイルを新しいリリースの配列データベース・ファイルに置き換えることにより配列データベースを更新することができます。MSDB、nr、EST、Swiss-Prot などの配列データベースのように、配列データベースが NCBI が管理している生物分類関連ファイルを利用している場合は、これらの更新ファイルも同時に更新する必要があります。

データベースの更新に必要なファイルを準備した後は、Mascot Monitor プログラムが自動的に動作し、配列データベースの構築、テスト検索などのプロセスを経て更新が完了します。更新プロセスを実行している最中でも MASCOT 検索を停止する必要はありません。また、MASCOT 検索が終了した後に新旧の配列データベースが入れ替わり、新しい配列データベースが有効になります。

標準でインストールされる db_update.pl プログラムを利用することにより、更新に必要なファイル群のダウンロードも含めて全自動で配列データベースの更新することができます。

2-1 Mascot Monitor プログラム

Mascot Monitor プログラムの実態は、

```
C:\inetpub\mascot\bin\ms-monitor.exe
```

です。Mascot Monitor プログラムは mascot.dat ファイルの Databases セクションで指定された current フォルダを常に監視しており、current フォルダ内に mascot.dat ファイルの中で指定された名称の配列データベース・ファイルが作成された時点で次の更新プロセスを開始します。

(1) 配列データベース・ファイルの圧縮

配列データベース・ファイルを構成する各エントリアクセション番号、説明、生物分類、配列に分割し、圧縮します。同時に、アクセション番号の重複、生物分類名のスペルのチェックを実行します。

(2) テスト検索

サンプルデータを使って MASCOT 検索を実施し、配列データベースの動作をチェックします。

(3) メモリへのマップ

mascot.dat の設定に従って、圧縮された配列データベース・ファイルをメモリ

にマップします。

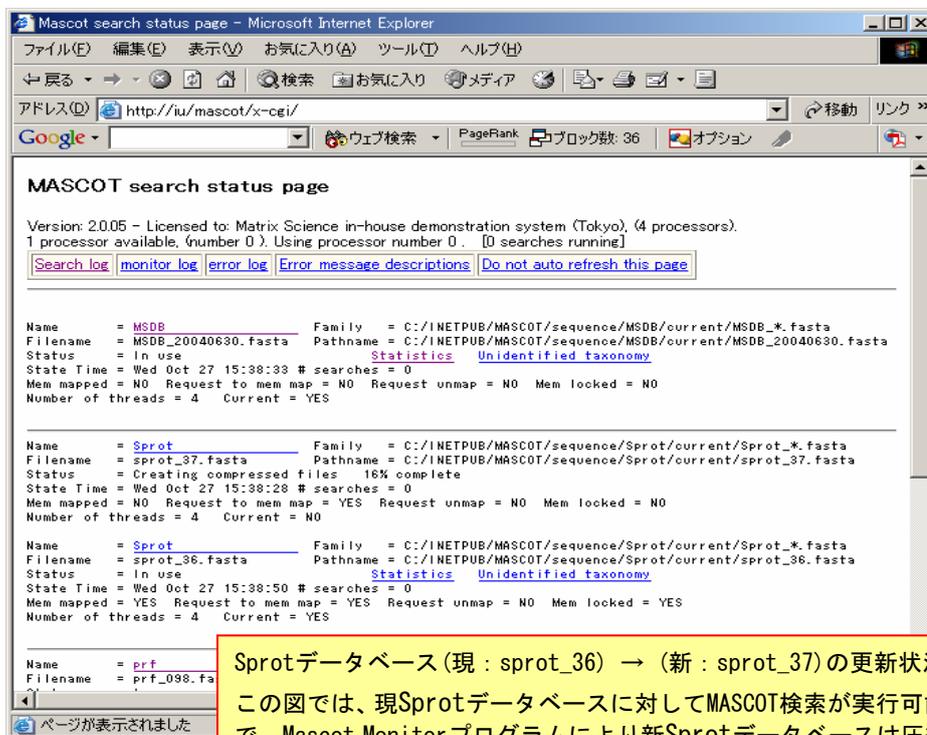
(4) 旧配列データベース・ファイルの移動

古い圧縮されたファイル群を削除し、古い FASTA ファイルを old フォルダに移動します。old フォルダに存在するより古い FASTA ファイルは削除されます。

Mascot Monitor プログラムの更新プロセスは、

<http://MASCOTホスト名/mascot/x-cgi/ms-status.exe>

画面の対応する配列データベース・ブロックで確認することができます。Sprot (Swiss-Prot データベース) の更新状況の例を示します。



Sprotデータベース (現 : sprt_36) → (新 : sprt_37) の更新状況

この図では、現Sprotデータベースに対してMASCOT検索が実行可能な状態 (Status = In use) で、Mascot Monitor プログラムにより新Sprotデータベースは圧縮プロセスが進行していません (Status = Creating compressed files 16% complete)。

status 行は、更新プロセスの進行とともに、

Status = Creating compressed files n% complete

Status = Running 1st test

Status = First test just run OK

Status = Trying to memory map files

Status = Just enabled memory mapping

Status = In use

のように変化し、Status = In use になった時点でデータベースの更新が完了します。

2-2 手動更新

更新に必要な配列データベース・ファイルを準備し、mascot.dat ファイルの Databases セクションで指定された current フォルダに、指定された名称の配列データベース・ファイルを作成してください。Mascot Monitor プログラムは current フォルダに配列データベース・ファイルが作成された時点で更新プロセスを開始します。

指定された配列データベース・ファイル名は current フォルダの中で設定してください。指定された配列データベース・ファイル名を持つファイルを他のフォルダから current フォルダにコピーまたは移動する場合、その処理が完了するまでに時間がかかりますので、その処理が完了する前に Mascot Monitor プログラム動作し、配列データベースのセットアップが正しく完了しない可能性があります。

NCBIInr データベースの手動更新の例を示します。NCBI のサイトから FASTA ファイル（圧縮ファイル名は nr.gz）と生物分類ファイル群（圧縮ファイル名は gi_taxid_prot.dmp.gz と taxdump.tar.gz）をダウンロードしてください。

NCBIInr データベースの手動による更新例

[1] 配列データベース・ファイルのダウンロード

“C:\inetpub\mascot\sequence\NCBIInr\incoming” フォルダを保存先として、次のファイルをダウンロードしてください。

<ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz>

[2] 生物分類(Taxonomy)関連圧縮ファイルのダウンロード

“C:\inetpub\mascot\taxonomy\temp” フォルダを保存先として、次の2つのファイルをダウンロードしてください。

ftp://ftp.ncbi.nih.gov/pub/taxonomy/gi_taxid_prot.dmp.gz

<ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz>

[3] 圧縮ファイル解凍

ダウンロードした圧縮ファイルを解凍してください。

[4] Taxonomyファイル群の移動

“C:\inetpub\mascot\taxonomy\temp” フォルダにある解凍済みファイルを “C:\inetpub\mascot\taxonomy” フォルダに移動してください（同名ファイルは上書き保存してください）。

[5] nrファイルの移動

“C:\inetpub\mascot\sequence\NCBI\nr\incoming” フォルダにあるnrを “C:\inetpub\mascot\sequence\NCBI\nr\current” フォルダに移動してください。また、nrファイルのプロパティから更新日を確認してください。

[6] 配列データベース・ファイル名の変更

nrファイルをNCBI_nr_20040808.fasta（更新日が2004年8月8日の場合）に名称変更してください。

[7] 更新プロセスの確認

<http://MASCOTホスト名/mascot/x-cgi/ms-status.exe> 画面のNCBI_nrブロックで配列データベース・ファイルの更新状況を確認してください。

Swiss-Prot データベースの手動更新の例を示します。Swiss-Prot データベースは、FASTA ファイル (uniprot_sprot.fasta.gz)、詳細なエントリ情報が記載されたリファレンス・ファイル (uniprot_sprot.dat.gz)、生物分類ファイル (speclist.txt) が組になっています。

Swiss-Prot データベースの手動による更新例

[1] 配列データベース・ファイル群のダウンロード

“C:\inetpub\mascot\sequence\Sprot\incoming” フォルダを保存先として、次の3つのファイルをダウンロードしてください。

<ftp://ftp.ebi.ac.uk/pub/databases/uniprot/knowledgebase/reldate.txt>

ftp://ftp.ebi.ac.uk/pub/databases/uniprot/knowledgebase/uniprot_sprot.fasta.gz

ftp://ftp.ebi.ac.uk/pub/databases/uniprot/knowledgebase/uniprot_sprot.dat.gz

[2] Taxonomy (生物分類) 関連圧縮ファイルのダウンロード

“C:\inetpub\mascot\taxonomy\temp” フォルダを保存先として、次の2つのファイルをダウンロードしてください。

<ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz>

<ftp://ftp.ebi.ac.uk/pub/databases/uniprot/knowledgebase/docs/speclist.txt>

[3] 解凍

ダウンロードした圧縮ファイルを解凍してください。

[4] Taxonomyファイル群の移動

“C:\inetpub\mascot\taxonomy\temp” フォルダにある解凍済みファイルを “C:\inetpub\mascot\taxonomy” フォルダに移動してください（同名ファイルは上書き保存してください）。

[5] 配列データベース・ファイル群の移動

“C:\inetpub\mascot\sequence\sprot\incoming” フォルダにある3つのファイルを “C:\inetpub\mascot\sequence\sprot\current” フォルダに移動してください。また、reldate.txtファイルを開き、リリース番号またはリリース日を確認してください。

[6] 配列データベース・ファイル名の変更

データベース・ファイルをリリース番号またはリリース日を使った次の名称に変更してください。

uniprot_sprot.fasta.gz → sprot_37.fasta
uniprot_sprot.dat.gz → sprot_37.dat

[7] 更新プロセスの確認

<http://MASCOTホスト名/mascot/x-cgi/ms-status.exe>画面のSprotブロックで配列データベース・ファイルの更新状況を確認してください。

MSDB、EST_xxxxx (xxxxx=human、mouse、others)、IPI_xxxxx (xxxxx=HUMAN、MOUSE、RAT) については次のファイル群をダウンロードしてください。

(1) MSDB

配列データベース・ファイル：

<ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.fasta.Z>

<ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.nam.Z>

<ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.ref.Z>

生物分類ファイル：

<ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz>

(2) EST_xxxxx (xxxxx=human、mouse、others)

配列データベース・ファイル：

ftp://ftp.ncbi.nih.gov/blast/db/FASTA/est_XXXXX.gz

生物分類ファイル：

ftp://ftp.ncbi.nih.gov/pub/taxonomy/gi_taxid_nucl.dmp.gz

<ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz>

(3) IPI_XXXXX (XXXXX=HUMAN、MOUSE、RAT)

配列データベース・ファイル：

<ftp://ftp.ebi.ac.uk/pub/databases/IPI/current/ipi.XXXX.fasta.gz>

生物分類ファイル：

なし。

2-3 自動更新

MASCOT Serverは、配列データベースを自動的に更新するためのプログラム db_update.pl をサポートしています（標準でインストールされます）。db_update.pl プログラムを動作させることにより、MASCOT検索に対して有効になっている配列データベースを最新のリリースに更新することができます。また、Windowsが標準でサポートしているタスクスケジューラやUnix/LinuxがサポートしているCronなどの、日時指定自動実行プログラムと組み合わせることにより、配列データベースを常に最新の状態に保つことができます。

db_update.plプログラムは次の機能を持っています。

- (1) 配列データベース関連ファイル（FASTA、リファレンスなど）の自動ダウンロード
- (2) Taxonomy定義ファイルのダウンロード
- (3) ワイルドカード（*）を利用したファイル名に対応
- (4) 圧縮ファイルの解凍、ファイル名の変更、ディレクトリの変更
- (5) ダウンロード最終履歴チェック機能
- (6) Passive FTP、HTTP/FTPプロキシ認証が可能

配列データベースを自動的に更新するためには次の4つのプログラム

(db_update.pl、gzip.exe、tar.exe、wget.exe) が必要です。

なお、弊社から出荷されたMASCOT Serverシステムには自動更新に必要なすべてのプログラムはインストールされていますのでこの操作は不要です。

db_update.pl :

MASCOT Serverのインストール時に、

C:\inetpub\mascot\bin\db_update.pl

ファイルとしてインストールされます。

gzip.exe、tar.exe、wget.exe :

次のサイトからダウンロードすることができます。

<http://www.gnu.org/software/gzip/gzip.html>

<http://www.gnu.org/software/tar/tar.html>

<http://www.gnu.org/software/wget/wget.html>

なお、MASCOT ServerのインストールCDROM (mskkフォルダの中のwbinフォルダ) にもこれらのプログラムが含まれています。

gzip.exe、tar.exe、wget.exeプログラムは、

C:\inetpub\mascot\binまたはC:\winnt\system32

(Windows XPの場合はC:\windows\system32です)

フォルダにコピーしてください (Windowsのシステムパスが有効な他のフォルダにコピーしても動作します)。

db_update.plプログラムは引数を指定してコマンドプロンプトから実行します。引数はダウンロードする配列データベースの種類に対応しており、次の引数を使用することができます。

引数	配列データベース名	ダウンロードサイト
MSDB_from_NCBI	MSDB	NCBI
MSDB_from_EBI	MSDB	EBI
NCBINr_from_NCBI	NCBINr	NCBI
EST_human_from_NCBI	EST_human	NCBI
EST_mouse_from_NCBI	EST_mouse	NCBI
EST_others_from_NCBI	EST_others	NCBI
Sprot_fasta_only_from_EBI	Swiss-Prot (FASTAファイル)	EBI
Sprot_complete_from_EBI	Swiss-Prot	EBI
Trembl_complete_from_EBI	TrEMBL	EBI
IPI_human_from_EBI	IPI_human	EBI
IPI_mouse_from_EBI	IPI_mouse	EBI

ここでは、NCBINrデータベースの自動更新を例にして、db_update.plプログラムの使用方法を説明します。

db_update.pl プログラムを利用した NCBI nr データベースの自動更新

[1] コマンドプロンプトの起動

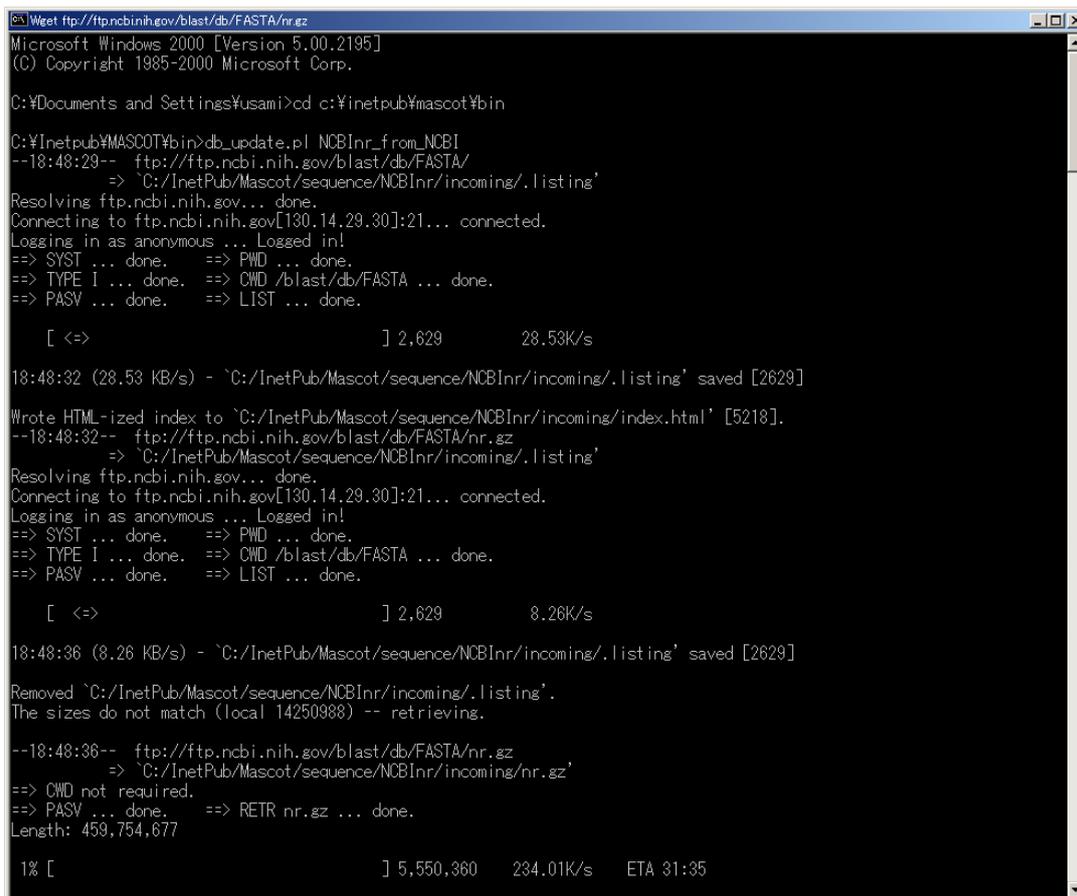
[スタート]メニュー → [プログラム] → [アクセサリ] → [コマンドプロンプト]を選択してください。コマンドプロンプトが起動します。

[2] db_update.pl プログラムの実行

ここでは、NCBIのサイトからNCBI nr データベース及びその関連ファイルをダウンロードしますので、キーボードから次のコマンドを入力し、リターン・キーを押してください。

```
cd C:\inetpub\mascot\bin
db_update.pl NCBI nr_from_NCBI
```

関連ファイル群のダウンロードが始まります。



```
Microsoft Windows [Version 5.00.2195]
(C) Copyright 1985-2000 Microsoft Corp.

C:\Documents and Settings\Yusami>cd c:\inetpub\mascot\bin
C:\inetpub\mascot\bin>db_update.pl NCBI nr_from_NCBI
--18:48:29-- ftp://ftp.ncbi.nih.gov/blast/db/FASTA/
=> `C:/InetPub/Mascot/sequence/NCBI nr/incoming/.listing'
Resolving ftp.ncbi.nih.gov... done.
Connecting to ftp.ncbi.nih.gov[130.14.29.30]:21... connected.
Logging in as anonymous ... Logged in!
==> SYST ... done. ==> PWD ... done.
==> TYPE I ... done. ==> CWD /blast/db/FASTA ... done.
==> PASV ... done. ==> LIST ... done.

[ <-> ] 2,629 28.53K/s

18:48:32 (28.53 KB/s) - `C:/InetPub/Mascot/sequence/NCBI nr/incoming/.listing' saved [2629]

Wrote HTML-indexed index to `C:/InetPub/Mascot/sequence/NCBI nr/incoming/index.html' [5218].
--18:48:32-- ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz
=> `C:/InetPub/Mascot/sequence/NCBI nr/incoming/.listing'
Resolving ftp.ncbi.nih.gov... done.
Connecting to ftp.ncbi.nih.gov[130.14.29.30]:21... connected.
Logging in as anonymous ... Logged in!
==> SYST ... done. ==> PWD ... done.
==> TYPE I ... done. ==> CWD /blast/db/FASTA ... done.
==> PASV ... done. ==> LIST ... done.

[ <-> ] 2,629 8.26K/s

18:48:36 (8.26 KB/s) - `C:/InetPub/Mascot/sequence/NCBI nr/incoming/.listing' saved [2629]

Removed `C:/InetPub/Mascot/sequence/NCBI nr/incoming/.listing'.
The sizes do not match (local 14250988) -- retrieving.

--18:48:36-- ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz
=> `C:/InetPub/Mascot/sequence/NCBI nr/incoming/nr.gz'
==> CWD not required.
==> PASV ... done. ==> RETR nr.gz ... done.
Length: 459,754,677

1% [ ] 5,550,360 234.01K/s ETA 31:35
```

コマンドプロンプトはdb_update.pl プログラムのすべての処理が完了した時点で自動的に閉じますので、強制的に閉じないでください。

[3] ログ

db_update.plプログラムによるNCBInrデータベースのダウンロードに関わるログは、

```
C:¥inetpub¥mascot¥logs¥ftp_log.log
```

に記録されます。また、前回のダウンロード履歴（nr.gzファイルの作成日とファイルサイズ）は、

```
C:¥inetpub¥mascot¥sequence¥NCBInr¥incoming¥.history
```

に記録され、この情報は配列データベースが更新されていない場合はダウンロードを中止するために利用されています。何らかの理由でダウンロードが完結しなかった場合でも、.historyファイルに前回のダウンロード履歴が残されますので、db_update.plプログラムを再実行した際に、

```
No update available for ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz
```

が表示され、プログラムが終了します。この場合は、.historyファイルを削除するか、履歴情報を編集した後に再度実行してください。

[4] 配列データベースのセットアップ

NCBInr データベースのファイル群のダウンロードが終了した時点で、Mascot Monitor プログラムが動作し、データベースのセットアップ・プロセスが進行します。

<http://MASCOTホスト名/mascot/x-cgi/ms-status.exe>

画面の NCBInr データベース・ブロックの Status 行で確認することができます。セットアップ・プロセスの進行とともに Status 行は、

```
Status = Creating compressed files    n% complete
Status = Running 1st test
Status = First test just run OK
Status = Trying to memory map files
Status = Just enabled memory mapping
Status = In use
```

のように変化し、Status = In use になった時点で配列データベースの更新が完了します。

2-4 Windows のタスクスケジューラの設定方法

Windowsのタスクスケジューラは、設定されたスケジュールに従って、指定したプログラムなどの処理をバックグラウンドで実行するためのユーティリティープログラムです。次の内容を設定することができます。

(1) タスクの定義

複数の独立したタスクを定義することができます。

(2) タスクの実行

指定日時（日/週/月間隔、コンピュータの起動時などを含む）にタスクを実行するように設定することができます。また、即座に実行することもできます。

(3) ユーザ制限

特定のユーザがタスクを実行できるよう、ユーザ制限を設定することができます。

(4) 1 回だけのタスク実行

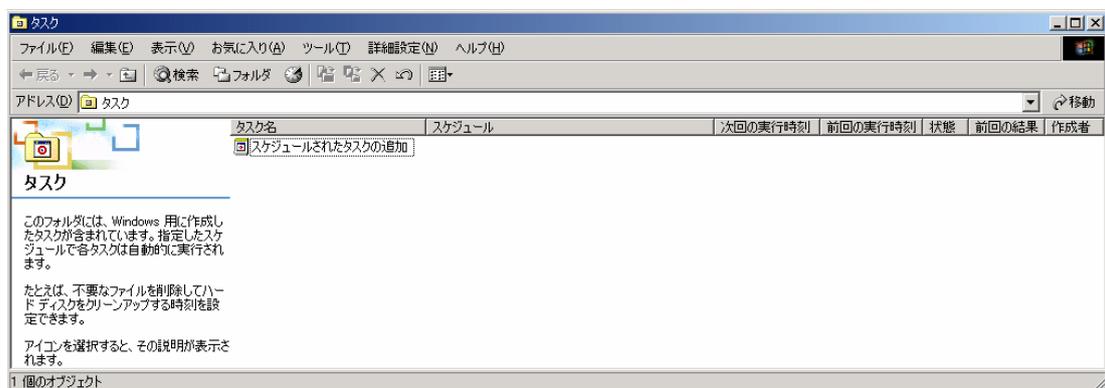
タスク終了後にそのタスク定義を削除することができます。

タスクスケジューラを利用して、データベース自動更新プログラムである db_update.pl を指定日時に自動的に実行することができます。ここでは、NCBIInr を自動更新するためのタスクスケジューラの設定例を説明します。

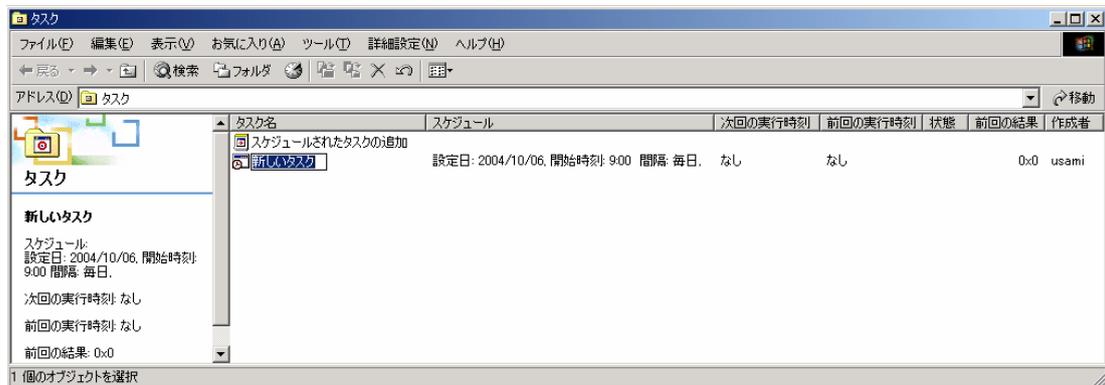
NCBIInr データベースを指定日時に自動更新するためのタスクスケジューラの設定方法

[1] 新規タスクの定義

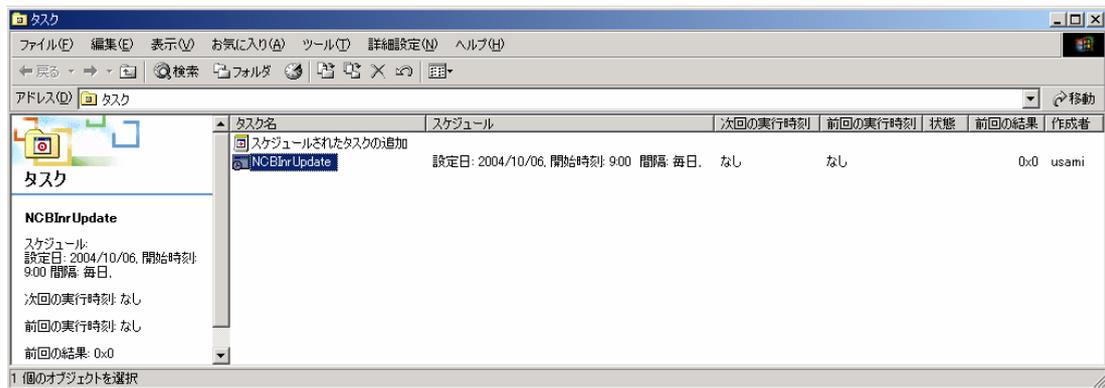
[スタート]メニュー → [プログラム] → [アクセサリ] → [システムツール] → [タスク] を選択してください。次のウインドが表示されます。



[ファイル] → [新規] → [タスク]を選択してください。[新しいタスク]が追加されます。

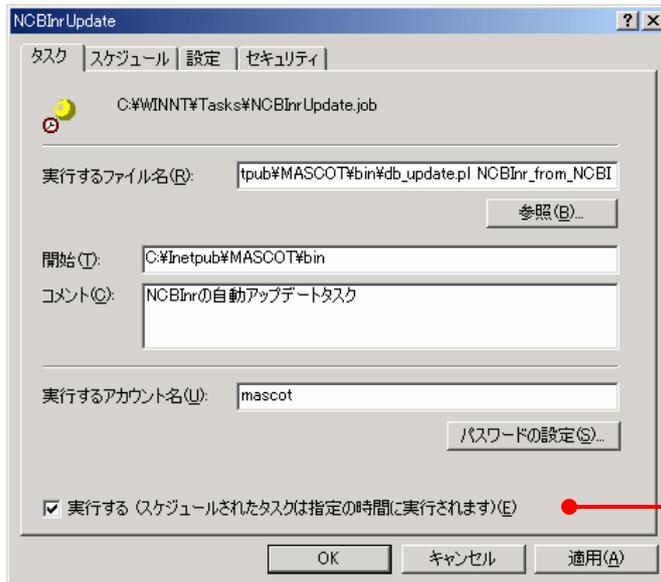


追加された[新しいタスク]の名称を[NCBInrUpdate]に変更してください。



[2] 実行するプログラムの定義

[NCBInrUpdate]タスクをダブルクリックしてください。プロパティ設定ウインドが開きます。



このチェックをはずすとこのタスクを無効にすることができます。

[タスク]タブの[実行するファイル名]入力欄に、

```
C:\perl\bin\perl.exe C:\Inetpub\MASCOT\bin\db_update.pl NCBIInr_from_NCBI
```

を入力してください。ここで入力した内容の意味は、

C:\perl\bin\perl.exe	:	perlプログラム実体
C:\Inetpub\MASCOT\bin\db_update.pl	:	db_update.plプログラムの実体
NCBIInr_from_NCBI	:	db_update.plプログラムの引数

になります。

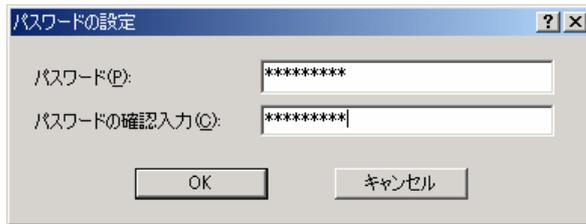
[開始]入力欄には、db_update.plプログラムを含むフォルダの存在場所である、

```
C:\Inetpub\MASCOT\bin
```

を入力してください。

[実行するアカウント名]入力欄にMASCOT Serverをインストールしたアカウント（通常はmascotです）を入力してください。

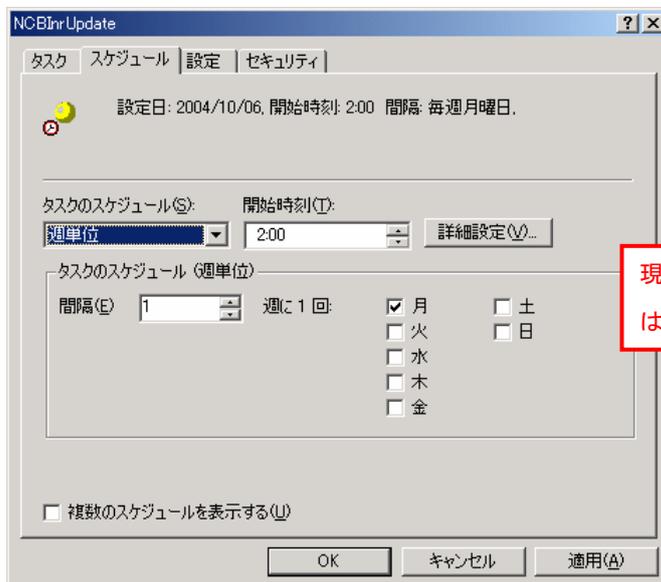
[パスワードの設定]ボタンを押すとパスワードの設定ウインドが表示されますので、パスワードを入力し、[OK]ボタンを押してください。



[適用]ボタンを押してください。

[3] スケジュールの設定

[スケジュール]タブの[タスクのスケジュール]および[開始時刻]入力欄に、このタスクを実行したい日時を設定し、[適用]ボタンを押してください。



[4] その他の設定

必要に応じて、[設定]タブ、[セキュリティ]タブで細かな条件を設定することができます。設定が終了したら、[OK]ボタンを押してプロパティ設定ウインドを閉じてください。

[5] タスクの即時実行

[3]で設定したスケジュールで[NCBInrUpdate]タスクは自動的に実行されますが、タスクを即時実行したい場合は、[NCBInrUpdate]タスクを右クリックして、[タスクの実行]を選択してください。コマンドプロンプトが開き、db_update.plプログラムによるNCBInrのダウンロードが始まります。なお、コマンドプロンプトは強制的に閉じないでください。db_update.plプログラムのすべての処理が完了した時点で自動的に閉じます。

2004-8-10現在、NCBInrデータベースの圧縮ファイルであるnr.gzのサイズは約400MBです。また、数10MB

のサイズを持つTaxonomy関係のファイル群も同時にダウンロードされますので、ネットワークの速度によりダウンロードには数時間要する場合があります。また、NCBIのサイトが混雑していたり、ネットワークトラフィックの状況変化などの理由により、ダウンロードが完結しない場合があります。この場合はコマンドプロンプトを閉じ、時間をおいてから再度[NCBIInrUpdate]タスクを実行してください。

[6] その他

db_update.plプログラムによるNCBIInrデータベースのダウンロードに関わるログは、

```
C:¥inetpub¥mascot¥logs¥ftp_log.log
```

に記録されます。また、前回のダウンロード履歴 (nr.gzファイルの作成日とファイルサイズ) は、

```
C:¥inetpub¥mascot¥sequence¥NCBIInr¥incoming¥.history
```

に記録され、この情報はデータベースが更新されていない場合はダウンロードを中止するために利用されています。従いまして、何らかの理由でダウンロードが完結せず、db_update.plプログラムを再実行した際に、

```
No update available for ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz
```

が表示されプログラムが終了した場合は、.historyにダウンロード履歴が残されていることがありますので、.historyファイルを削除するか、履歴情報を編集した後に再度実行してください。

3 新規配列データベースのセットアップ

新規配列データベースの登録とセットアップの大きな流れは次のようになります。

- (1) 新規配列データベースの FASTA ファイルの作成
- (2) エントリの書式の確認と情報抽出のための正規表現の検討
- (3) 新規データベース格納フォルダの作成
- (4) 設定ファイル (mascot.dat) への新規データベースの仮登録
- (5) 仮登録した新規配列データベースに対する動作確認
- (6) 新規配列データベースの本登録
- (7) 検索テスト

3-1 新規配列データベースのFASTA ファイルの作成

新規配列データベースをFASTA形式の配列データベース・ファイルとして作成します。
FASTA形式の配列データベースは、

- 1行目：コメント行（1文字目は >、> 以降はアクセッション番号を含む文字列）
- 2行目：配列行（改行および複数行構成可、空白文字は不可）

の内容がひとつの単位（エントリ）となり、この単位（エントリ）が複数個結合されています。

ここでは、**NEW** という名称を持つ配列データベースを例にして説明を進めます。**NEW** データベースは次のような10個のエントリを持ち、ファイル名は **NEW_20040101.fasta** とします。なお、MASCOT Serverは大文字と小文字を区別しますのでご注意ください。

NEW_20040101.fastaデータベース・ファイルの内容

```
>00001 Test description for a protein aaa
MIVASKLPSLPLPFSFVFSILLFVFEMMVVLHF
>00002 Test description for a protein bbb
MASLTDLKVNLNLSDTATEKIIAEYMIWIGSGMDLRRKRPPV
>00003 Test description for a protein ccc
MAKFKNL GKSPKRVAKEEIILLDKGLFASCSFTDLGLHP
>00004 Test description for a protein ddd
MIVASKLPSLPLPFSFVFSILPROTEINDDDL FVFEMMVVLHF
>00005 Test description for a protein eee
MIVASKLPSLPLPFSFAAADKRTMSVFSILPROTEINEEELFVFEMMVVLHF
>00006 Test description for a protein fff
MAMKRIVASKLPSLTPIFSFAAADKRTMSVFSILPROTEINFFFLFVFEMMVVLHF
>00007 Test description for a protein ggg
MARPPSSGAQQGAKPGADKNLRFLLQKVLKQSDVGALGRIVLPKEAEGISIPMEDI GTSRVWSMRYRF
>00008 Test description for a protein hhh
MLIVQFEPDSPDHTWAQATDNGSIKNHEPHGNSPHYRKLGIKWL MNTMYELTVDKRLERLLGGGR IYK
>00009 Test description for a protein iii
MLIVQFEPDSPDHTWAQATDNGSIKNHEPHGNTLHYRKLGIKWL MNTMYELTVDKRLERLGRIPLYYK
>ACC00010 Test description for a protein jjj
AAGRGRGRCRGS DPAAGDDVLDLIDELLN FATLSDSMPWEDEPLFPDDVGMMIEDAISHQPHRG
```

通常、配列データベースは中身の修正やエントリの追加などの変更が加えられ、新しいバージョンまたはリリースが作成されますので、配列データベース・ファイル名は、(配列データベース名)と(新規バージョンまたはリリース日などの情報)をアンダーバー”_”で結合した形式にしてください。

3-2 エントリの書式の確認と情報抽出のための正規表現の検討

NEW データベースの各エントリを構成するAccession番号(00001など)とDescription(Test description for a protein aaaなど)をMASCOT Serverに認識させ、MASCOT Serverがこれらの文字列を正しく抽出し、利用できるようにするために、正規表現を使ってこれらの文字列を定義します。

何通りかの適当な正規表現が考えられますが、エントリの1行目からAccession番号に対応する文字列を抽出するために、たとえば次の正規表現を適用することができます。

```
>¥([ ^ ]*¥)
```

この正規表現は、(>) + (スペース以外の任意個の文字)で構成された文字列に含まれる(スペース以外の任意個)の文字列部分をAccession番号として抽出する、という意味になります。なお、正規表現で定義した文字列全体からある文字列部分を抽出し、その文字列部分をMASCOTに認識させるために、その文字列部分をカッコ ¥(と ¥)で囲みます。ここで、前カッコと後カッコの前に ¥ (バックスラッシュ)がついていますが、正規表現ではカッコは特別な意味を持つため、バックスラッシュ¥でその意味を無効にしています。

同様に、Descriptionの抽出に対しては次の正規表現を適用することができます。

```
>[ ^ ]* ¥(. *¥)
```

この正規表現は、(>) + (スペース以外の任意個の文字) + (スペース) + (任意の任意個の文字)で構成された文字列に含まれる(任意の任意個の文字)の文字列をDescriptionとして抽出する、という意味になります。

3-3 新規配列データベース格納フォルダの作成

MASCOT Server はひとつの配列データベースに対して3つのフォルダを必要とします。

```
C:¥inetpub¥mascot¥sequence¥配列データベース名¥current
C:¥inetpub¥mascot¥sequence¥配列データベース名¥incoming
C:¥inetpub¥mascot¥sequence¥配列データベース名¥old
```

次の操作でNEW データベースに必要なフォルダを作成し、NEW_20040101.fasta ファイルを **current** フォルダにコピーしてください。

NEW_20040101.fastaデータベース・ファイルの内容

[1] MASCOT Serverにログオン

mascotユーザでログオンしてください。

[2] フォルダの作成

次の3つのフォルダを作成してください。

```
C:¥inetpub¥mascot¥sequence¥NEW¥current
C:¥inetpub¥mascot¥sequence¥NEW¥incoming
C:¥inetpub¥mascot¥sequence¥NEW¥old
```

[3] NEW_20040101.fastaファイルのコピー

C:¥inetpub¥mascot¥sequence¥NEW¥currentフォルダにNEW_20040101.fastaファイルをコピーしてください。また、NEW_20040101.fastaファイルのセキュリティがEveryone/フルコントロールになっていることを確認してください。

3-4 設定ファイル (mascot.dat) への新規配列データベースの仮登録

配列データベースを新規に登録しセットアップするために、mascot.datファイルのDatabasesセクション、PARSEセクション及びWWWセクションに必要な情報を追加します。次の操作で NEW データベースの仮登録を行ってください。

NEWデータベースの仮登録

[1] MASCOT Serverにログオン

mascotユーザでログオンしてください。

[2] mascot.datファイルの編集準備

mascot.datファイルを別名でコピーし（たとえば、mascot.dat.backup）、バックアップしてください。
テキストエディタを起動し、mascot.datファイルを開いてください。

[3] PARSEセクションの編集

NEW データベースの各エントリを構成するAccession番号、DescriptionをMASCOT Serverに認識させるために、Accession番号およびDescriptionに対応する文字列を抜き出すための正規表現ルールを追加します。
【2】で検討した正規表現ルールを、ルール番号が連番になるように（この例では28と29を使用）追加してください。追加したルール番号はDatabasesセクションで使用します。なお、正規表現ルール部分は“ ”で囲んでください。

```
# Parse rule for NEW Database
RULE_28 ">¥([ ^ ]*¥)"
RULE_29 ">[ ^ ]* ¥(. *¥)"
```

[4] Databasesセクションの編集

NEW データベースの属性を、14個のパラメータで構成された1行で定義します。各パラメータはスペースまたはタブで区切ります。MSDBデータベースの例を示します。

```
MSDB C:/INETPUB/MASCOT/sequence/MSDB/current/MSDB_*.fasta AA 1 1 1 1 1 0 1 17 18 19 7
```

1. unique database name

MSDB：データベース名、他と区別できる唯一の名称

2. path to local FASTA file. Must use forward slashes and include wild card

C:/inetpub/mascot/sequence/MSDB/current/MSDB_*.fasta：配列データベースファイルの位置のフルパス表現、リリース番号等を代表するワイルドカード（*）が必要

3. AA for amino acid, NA for nucleic acid

AA：アミノ酸配列AA、塩基配列NA

4. approximate numbers of entries (not used)

1：未使用

5. unique integer ID (not used)

1：未使用

6. 1 for memory map, 0 otherwise

1：メモリマップ・スイッチ（0=オフ、1=オン）

7. number of blocks

- 1 : 1を使用
8. number of threads
1 : MASCOTライセンス数 (2ライセンスの場合は2、4ライセンスの場合は4など)
9. 1 for lock in memory, 0 otherwise
1 : メモリロック・スイッチ (0=オフ、1=オン)
10. 1 for local reference file, 0 otherwise
1 : ローカルなリファレンスファイルの存在表示 (0=存在しない、1=存在する)
11. index of parse rule for accession number from FASTA
17 : Accession番号の正規表現ルール番号
12. index of parse rule for description from FASTA
18 : Descriptionの正規表現ルール番号
13. index of parse rule for accession number from local reference file
19 : ローカルファイルからのAccession番号の正規表現ルール番号
14. index of taxonomy block, 0 if no taxonomy
7 : Taxonomy抽出ルール番号

Databaseセクションの最終行 (end行の前行) に **NEW** データベースに関する次の行を追加してください。

```
# NEW C:/INETPUB/MASCOT/sequence/NEW/current/NEW_*.fasta AA 1 1 1 1 1 0 0 28 29 0 0
```

なお、仮登録とするために、行頭にはコメント#が存在することに注意してください。

[5] WWWセクション

指定したAccession番号に対応する配列情報 (配列、Accession番号、Description等) を抜き出すためのプログラム (ms-getseq.exe) を利用するための設定です。WWWセクションの最終行 (end行の前行) に次の行を追加してください。

```
NEW_SEQ "8" "localhost" "80" "C:/INETPUB/MASCOT/x-cgi/ms-getseq.exe NEW #ACCESSION# seq"
```

各セクションの設定終了後、mascot.datファイルを上書き保存してください。

配列データベースが塩基配列で構成されている場合は、6通りの読枠 (six reading frames) が存在しますので、次の例のようにヒットした読枠の番号を #FRAME# を使って指定する必要があります。

```
NEW_SEQ "8" "localhost" "80" "C:/INETPUB/MASCOT/x-cgi/ms-getseq.exe NEW #ACCESSION# seq #FRAME#"
```

3-5 仮登録した新規配列データベースに対する動作確認

MASCOT Serverは、【4】で設定した内容が正しく動作するかどうかを判定するためのプログラムをサポートしており、ブラウザから利用することができます。

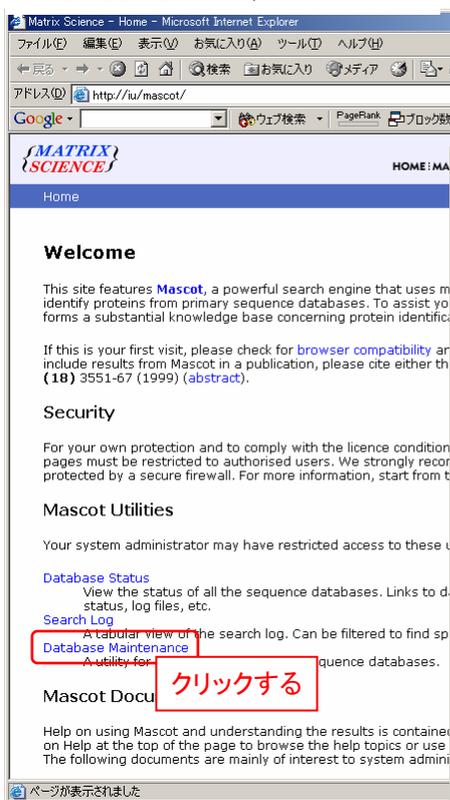
ブラウザを起動し、 URLとして、

<http://MASCOTホスト名/mascot/>

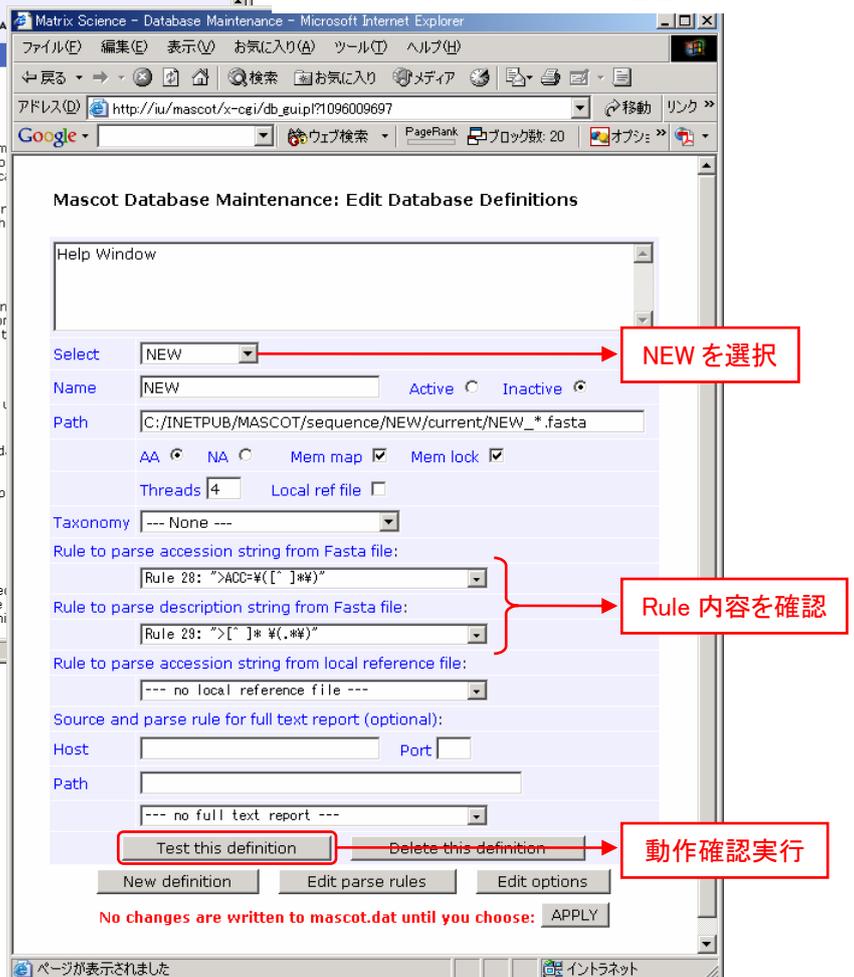
を指定してください。MASCOT Server の **トップページ** が表示されます。

トップページ 画面の中の中央より少し下にある **Database Maintenance** リンクをクリックしてください。**Mascot Database Maintenance** 画面が表示されます。

MASCOT Server のトップページ



Mascot Database Maintenance 画面



表示された **Mascot Database Maintenance** 画面において、次の操作を行ってください。

NEWデータベースの動作確認

[1] NEW データベースの選択

Select 項から「NEW」を選択してください。

[2] 正規表現ルールの確認

Rule to parse accession 及び **description string from Fasta file** 部分に表示されている正規表現ルール内容及び番号が正しいことを確認してください。

[3] 正規表現テスト結果の確認

Test this definition ボタンを押し、表示された動作テスト結果を確認してください。mascot.dat に追加した設定が正しく動作するかどうかを試験され、エラーがなければ次のような画面が表示されます。

動作テスト結果ページ

The screenshot shows a web browser window titled "Matrix Science - Database Maintenance - Microsoft Internet Explorer". The address bar shows the URL "http://iu/mascot/x-cgi/db_gui.pl?1096009901". The main content area displays the following information:

Mascot Database Maintenance

Testing Database Definition NEW

Warning: Definition is not active

Testing entries at beginning and end of
C:/INETPUB/MASCOT/sequence/NEW/current/NEW_20040101.fasta:

Accession	Description
00001	Test description for a protein aaa
00002	Test description for a protein bbb
00003	Test description for a protein ccc
00004	Test description for a protein ddd
00005	Test description for a protein eee
00006	Test description for a protein fff
00007	Test description for a protein ggg
00008	Test description for a protein hhh
00009	Test description for a protein iii
00010	Test description for a protein jjj

At the bottom of the page, there is a button labeled "Return to database definitions".

画面中の **Accession** の欄には、正規表現ルール番号28によって抽出されたAccession番号がリストされます。同様に、**Description** の欄には、正規表現ルール番号29によって抽出されたDescriptionの内容がリストされます。ここで何らかのエラーが生じた場合は、その内容が赤文字で出力されます。その場合は、

[4] エラー内容の確認

エラーの内容を確認してブラウザを閉じてください（**ブラウザは必ず閉じてください**）。

[5] mascot.dat の設定内容の再確認

mascot.dat の設定内容を再度確認し、正しく編集した上で、上書き保存してください。

[6] 再テスト

ブラウザを開き、再度[1]からやり直し、エラー出力がなくなるまで繰り返してください。

3-6 新規配列データベースの本登録

テキストエディタで mascot.dat ファイルを開き、Databases セクションの **NEW** データベースの登録行のコメント **#** をはずし、上書き保存してください。**NEW** データベースのセットアップが始まります。

NEWデータベースの本登録

(変更前)

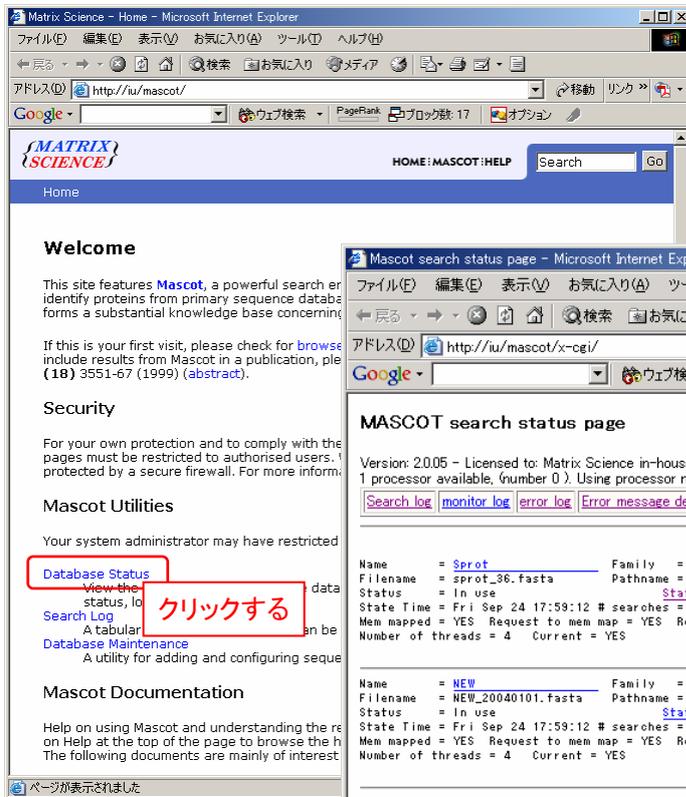
```
# NEW C:/INETPUB/MASCOT/sequence/NEW/current/NEW_*. fasta AA 1 1 1 1 1 0 0 28 29 0 0
```

(変更後)

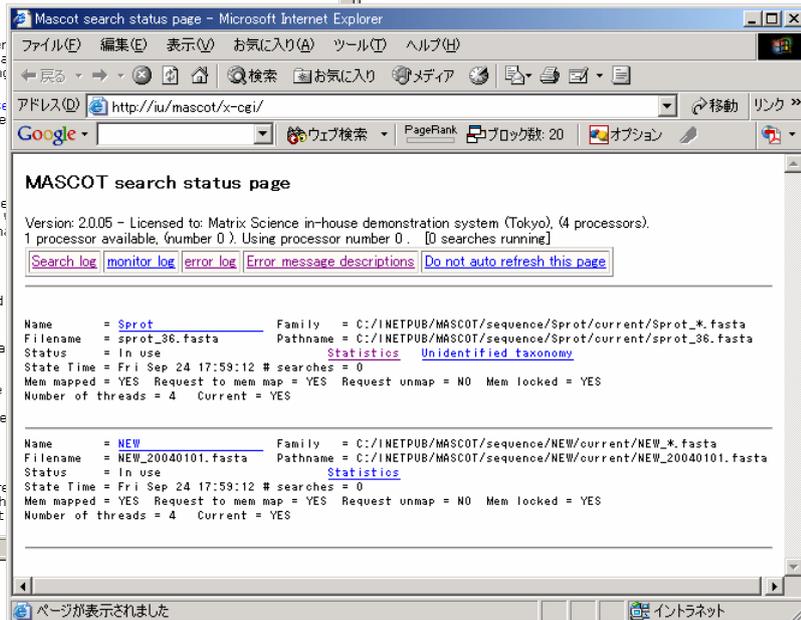
```
NEW C:/INETPUB/MASCOT/sequence/NEW/current/NEW_*. fasta AA 1 1 1 1 1 0 0 28 29 0 0
```

ブラウザを起動し、MASCOT のトップページ画面にある **Database Status** のリンクをクリックし、**MASCOT search status page** 画面を表示させてください。**NEW** データベースのセットアップ状況を確認することができます。

MASCOT Server のトップページ



MASCOT search status page



NEW データベースのセットアップの進行に伴い、NEW データベース・ブロック内の Status 行の表示が次のように変化します。

- Status = Creating compressed files n% complete
- Status = Running 1st test
- Status = First test just run OK
- Status = Trying to memory map files
- Status = Just enabled memory mapping
- Status = In use

In Use になった時点で、NEW データベースのセットアップが完了します。

3-7 検索テスト

ブラウザを起動し、検索条件設定画面を表示させてください。**Database** 項から **NEW** データベースを選択することができ、適当な質量データを使ってMASCOT検索が実行できることを確認してください。

Matrix Science - Mascot - MS/MS Ions Search - Microsoft Internet Explorer

HOME: MASCOT: HELP Search Go

Mascot > MS/MS Ions Search

MASCOT MS/MS Ions Search

Your name	test	Email	
Search title	search testing for NEW database		
Database	NEW		
Taxonomy	All entries		
Enzyme	Trypsin	Allow up to	1 missed cleavages
Fixed modifications	AB_old_ICATd0 (C) AB_old_ICATd8 (C) Acetyl (K) Acetyl (N-term) Amide (C-term)	Variable modifications	AB_old_ICATd0 (C) AB_old_ICATd8 (C) Acetyl (K) Acetyl (N-term) Amide (C-term)
Protein mass	kDa	ICAT	<input type="checkbox"/>
Peptide tol. ±	1.2 Da	MS/MS tol. ±	0.6 Da
Peptide charge	1+	Monoisotopic	<input checked="" type="radio"/> Average <input type="radio"/>
Data file		参照...	
Data format	Mascot generic	Precursor	m/z
Instrument	Default		
Overview	<input type="checkbox"/>	Report top	20 hits

Start Search ... Reset Form

Copyright © 2003 Matrix Science Ltd. All Rights Reserved.

以上で新規配列データベースの登録とセットアップは完了です。