

Sequence Database Download & Configuration

ASMS 2003



This talk will be mainly of interest to those people who administer an in-house Mascot server.

General procedure for setting up a new database

- **Choose a name**
 - Used in the search form drop down list, so needs to be short and descriptive.
- **Create a directory structure**
 - incoming, current, old
- **Find and download the files**
- **Update the Mascot configuration**
- **Bring on-line**
- **Set up automatic updating**

ASMS 2003



So, you've installed a copy of Mascot, which includes a copy of the MSDB protein database, and everything is working smoothly. Then you, or one of your colleagues, wants to search a different database. How do you go about adding a new database?

The general procedure is shown here:

<read from slide>

Let's focus on these last four steps in a little more detail

Download the files

- **Best done manually, the first time**
 - Verify file names and URL's
 - Do the files need unpacking or uncompressing?
- **Always require a FASTA file**
 - May require taxonomy
 - May require full text Ref / Dat file

Tip: Use full text file for Swiss-Prot and MSDB to get accurate taxonomy.

ASMS 2003

**MATRIX
SCIENCE**

First, catch your database, as Mrs Beeton would have said.

We'll talk later about automated downloading. But, when setting up a new database, you need to do everything manually at least once, to verify the filenames and URL's. If your running Mascot under Windows, you may find that you need to obtain utilities to decompress or untar the downloaded files.

At a minimum, we need a FASTA file. Some databases, such as Swiss-Prot and MSDB, also come with 'full text' files, containing comprehensive annotations. If a full text file is available and supported by Mascot, its a good idea to get it, if only because it will contain much more comprehensive taxonomy than can be included in a FASTA

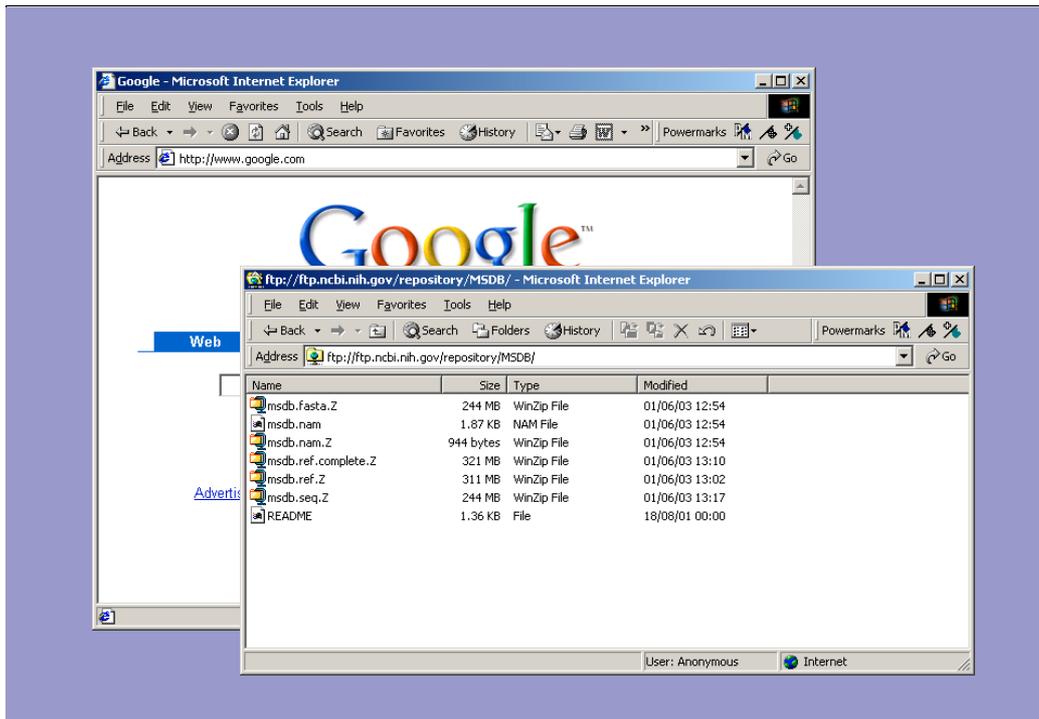
The slide displays two web browser windows and a collection of organism logos. The left window, titled "The NCBI ftp site - Microsoft Internet Explorer", shows the NCBI FTP site with a search bar and navigation links. The right window, titled "EBI Databases - Microsoft Internet Explorer", shows the EBI Databases website with a search bar and a list of databases. The background features logos for various organisms: UniProt, C.elegans, Zebrafish, Mosquito, Rat, Human, EMBL, SwissProt, Fly, Fugu, Mouse, Zebrafish, and TREMBL.

ASMS 2003 **MATRIX SCIENCE**

There are a few sites that provide comprehensive database repositories. Two of the best known are NCBI and EBI.

Here, you can download nr, GenBank, Swiss-Prot, EMBL, Trembl, etc.

For specialised databases, such as individual genomes, you may have to track down the FTP site of the group that compiles the database



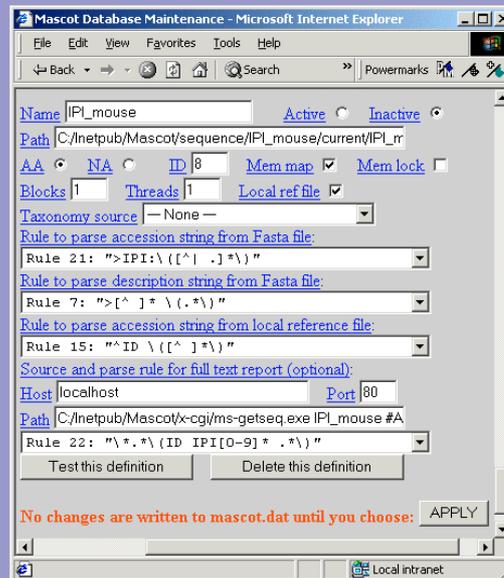
ASMS 2003



As with all web searches, a search engine such as Google is a good starting point. Once you find the database, you may have to study the README to figure out which files you need

Configuration

- Database maintenance utility
 - Browser based
- Safer than editing mascot.dat
- Always creates a backup
- Simple test facility



ASMS 2003



So, having caught your database, you need to tell Mascot how to cook it.

We recommend that you use the browser based, database maintenance utility, rather than edit the configuration file directly. This reduces the chance of typos and ensures that there is always a backup in case of problems.

It also has a simple test facility, that will pick up most of the common configuration problems

FASTA Format

```
>Title text
SEQUENCESEQUENCESEQUENCESEQUENCESEQUENCESEQUENCE
SEQUENCESEQUENCESEQUENCESEQUENCESEQUENCESEQUENCE
SEQUENCESEQUENCESEQUENCE
>Next title
NEXTSEQUENCE ..

>gi|6|bgi|Contig1.seq_7|2412 3299 [+3 L= 888] [Delayed
>20021010.2.1 1112073P09.y1 1112091F10.y1 1112073F0
>IPT:IPT00140097.1|REFSEQ_XP:XP_168061 Tax_Id=9606
>CCRB cytochrome c [validated] - rabbit
>gi|129249|sp|P02820|OSTC_BOVIN Osteocalcin precursor
>*ORF5 | start 2178-1309 | frame -1 | length=870 |
```

ASMS 2003



Perhaps this is a good moment to clarify exactly what we mean when we talk about a FASTA file.

FASTA is a very popular format. In a way, its like HTML ... it has become a universal standard because it is both very simple and very forgiving. On the downside, it isn't much of a standard ... almost anything goes.

FASTA specifies that there will be a title line, starting with a 'greater than' character, followed by one or more lines containing the sequence in 1 letter code.

The problem is the lack of a well defined syntax for the title line. Here are a handful of examples of FASTA title lines. As you can see, there isn't much similarity. It is essential that we can find a short, unique identifier or accession string for each sequence. But, as you can see from these examples, the position of the identifier and the delimiters (e.g. spaces, pipe symbols, commas) varies considerably

Parse Rules

- Parse rules are Basic Regular Expressions

```
>IPI:IPI00043251.2|REFSEQ_XP:XP_064505 Tax_Id=9606  
similar to keratin 18, cytoskeletal - human (fragment)
```

Accession from Fasta title: `">IPI:\([^| .]*\)"`

Description from Fasta title: `">[^]* \(.*\)"`

ASMS 2003



The way Mascot deals with this is to allow you to define the syntax for any particular database using a regular expression. Regular expression will be very familiar to anyone with a Unix background, but there may be a bit of a learning curve for someone with more of a Windows or Mac background.

Here, for example, we have a title line from the IPI human database. Lets say that we want to use IPI00043251 as the unique accession string and everything after the space should be treated as the description.

The regular expressions, or parse rules, used to extract this information look like this.

The string we want to extract is always within the backslashed parentheses. For the accession, we show the first few characters as literal text. We then say that we want to take all the following characters, stopping when we hit either a pipe symbol, a space, or a period. In fact, it is the period which applies in this example.

For the description, we discard everything up to and including the first space. Then, we take everything to the end of the title.

The screenshot shows the 'Database Maintenance' utility in a Microsoft Internet Explorer browser window. The interface includes fields for 'Path', 'AA', 'NA', 'ID', 'Mem map', 'Main lock', 'Blocks', 'Threads', 'Local ref file', 'Taxonomy source', and several 'Rule' fields for parsing accession and description strings. A 'Test this definition' button is visible at the bottom.

Callouts from the image:

- Include a timestamp or release number in the filename, e.g. Sprot_41.9.fasta** (points to the Path field)
- Full text DAT file is available for Swiss-Prot** (points to the Taxonomy source dropdown)
- TEST!** (points to the Test this definition button)
- DAT file gives very accurate taxonomy** (points to the Rule fields)

ASMS 2003



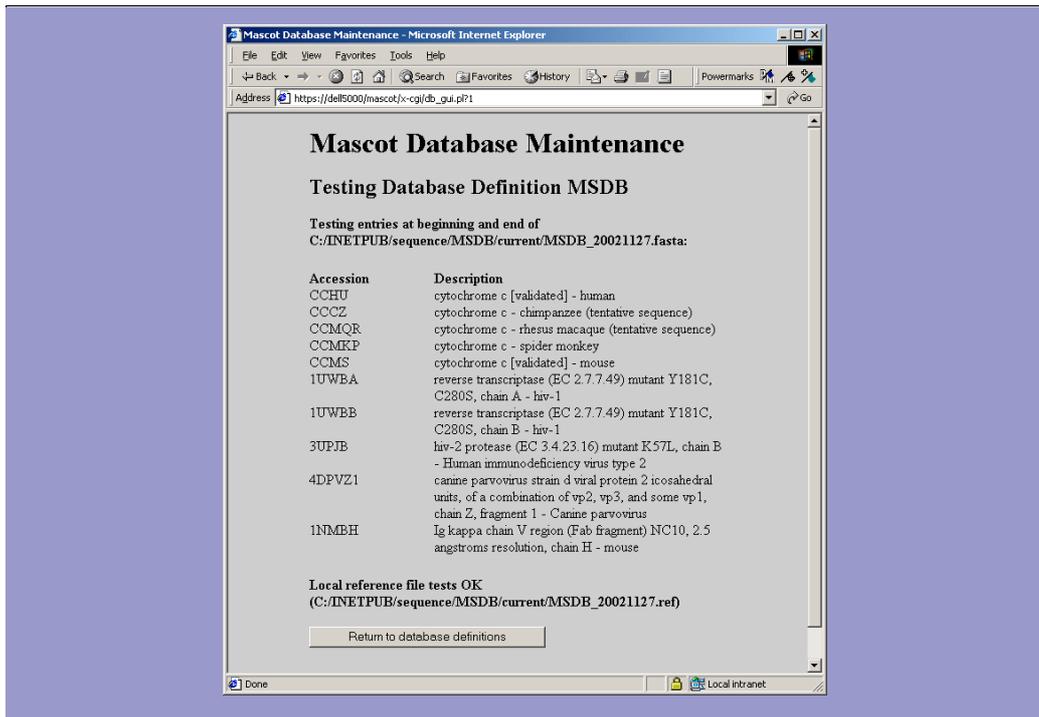
Deciding which parse rule to use, maybe even having to create a new parse rule, is definitely the most complicated part of configuring a new database. However, if we didn't use a tool like regular expressions, you would always be witing on us to add support into Mascot for new and unusual title formats.

The rest of the fields in database maintenance utility are pretty straight forward.

Its a good idea to include a version stamp or data stamp in the filename, so that you know what version of a database was used to obtain a particular result

In this example, Swiss-Prot, a full text DAT file is available, so we should make use of it to get accurate taxonomy

Once all the fields have been completed, hit the test button...



ASMS 2003



If you're unlucky, you'll see something like this. Looks like our first attempt at a parse rule wasn't quite right. OK, go back and try again
 Success

Bring database on-line

- **Monitor progress on the Database Status page**
 - Errors may appear in Monitor log or Error log
 - Fix the problem and retry

```
Name      = IPI\_human Family      = C:/INETPUB/sequence/IPI_human/current/IPI_human_*.fasta
Filename  = IPI_human_2.8.fasta Pathname = C:/INETPUB/sequence/IPI_human/current/IPI_human_2.8.fasta
Status    = In use      Statistics Unidentified taxonomy
State Time = Wed Jun 04 11:09:57 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = 1 Current = YES
```

ASMS 2003



Once the database tests successfully, we choose to apply the changes and bring the new database on-line

The test facility can't pick up every problem. In particular, it only tests a few of the entries at the start of the database and a few at the end. So, it's perfectly possible that something may go wrong in the middle.

If so, check the error log to see what the problem is, fix it, and re-try

Automatic updating with the database update script

- Download the fasta file plus all associated files, e.g. full text ref file, taxonomy indexes, Unigene
- Handles variable filenames via wild cards
- Takes care of uncompressing, unpacking, renaming and moving the files
- Time or version stamping; downloading a file only if a new one is available; resuming an interrupted download; passive FTP through a firewall; etc., etc.

ASMS 2003



Now, our new database is ready for searching. What about updates? We don't want to have to remember to download updates manually every day or every week. Mascot includes a script, the database update script, that will
<read from slide>

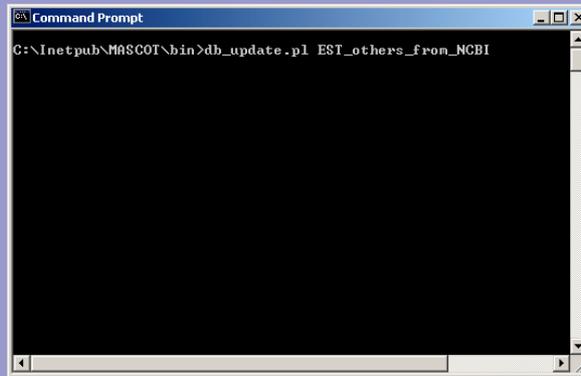
```
UltraEdit-32 - [C:\inetpub\MASCOT\bin\db_update.pl]
File Edit Search Project View Format Column Macro Advanced Window Help
db_update.pl
41 #####
42 # Safest to treat all path and URL information as case sensitive
43 # Windows paths MUST use forward slashes, not back slashes
44 # Windows paths CANNOT contain embedded spaces
45 # Windows paths should include drive letter, but UNC paths are not supported
46 #
47 if ($ENV{'WINDIR'}){
48 # Windows paths to Mascot root directory and utilities
49 # Must provide fully qualified path if utility not on the search path
50 $MASCOT = "C:/InetPub/Mascot";
51 $gzip = "gzip.exe";
52 $tar = "tar.exe";
53 $wget = "wget.exe";
54 $fmerge = "$MASCOT/bin/fmerge.exe"; # optional
55 $varsplic = "$MASCOT/bin/varsplic.pl"; # optional
56 } else {
57 # Unix paths to Mascot root directory and utilities
58 # Must provide absolute path if utility not on the search path
59 $MASCOT = "/usr/local/mascot";
60 $gzip = "gzip";
61 $tar = "tar";
62 $wget = "wget";
63 $fmerge = "$MASCOT/bin/fmerge"; # optional
64 $varsplic = "$MASCOT/bin/varsplic.pl"; # optional
65 }
66 #
67 # Other paths
68 #
69 $local_log_file = "$MASCOT/logs/ftp_log.txt";
70 $local_taxonomy_directory = "$MASCOT/taxonomy";
71 #
72 # Uncomment following path to enable UniGene index download
73 $local_unigene_directory = "$MASCOT/unigene";
74 #
75 # To fine tune wget, options can be added or modified below.
76 # If an option requires no value, assign an empty string.
77 # See wget documentation for further details.
78 # For example, if HTTP proxy server authentication is necessary:
```

The database update script is a perl script. If you've installed Mascot into the default path, you should be able to use this script as is to download updates for many widely used databases.

```
UltraEdit-32 - [C:\inetpub\MASCOT\bin\db_update.pl]
File Edit Search Project View Format Column Macro Advanced Window Help
db_update.pl
89 # Enter remote site URL's and local database directory paths below
90 #
91 # Variable filenames can be handled by using the wildcard characters * and ? in
92 # the FTP download URL. Wildcards cannot be used for HTTP download. Attempting
93 # to download multiple files with a single URL will download one file at most.
94 #
95 # Timestamping, to avoid downloading the same file on multiple occasions,
96 # will only work for FTP.
97 #
98 # MSDB_from_NCBI
99 #
100 if ($ARGV[0] eq "MSDB_from_NCBI") {
101     $db_name = "MSDB";
102     $local_incoming_directory = "$MASCOT/sequence/MSDB/incoming";
103     $local_current_directory = "$MASCOT/sequence/MSDB/current";
104     $fasta_file_url = "ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.fasta.Z";
105     $reference_file_url = "ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.ref.Z";
106     $name_file_url = "ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.nam.Z";
107     $taxonomy_file_url[0] = "ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz";
108     $version_regex = 'MSDB\s+Release\s+(\d+) ';
109 #
110 # MSDB_from_EBI
111 #
112 } elsif ($ARGV[0] eq "MSDB_from_EBI") {
113     $db_name = "MSDB";
114     $local_incoming_directory = "$MASCOT/sequence/MSDB/incoming";
115     $local_current_directory = "$MASCOT/sequence/MSDB/current";
116     $fasta_file_url = "ftp://ftp.ebi.ac.uk/pub/databases/MassSpecDB/msdb.fasta.Z";
117     $reference_file_url = "ftp://ftp.ebi.ac.uk/pub/databases/MassSpecDB/msdb.ref.Z";
118     $name_file_url = "ftp://ftp.ebi.ac.uk/pub/databases/MassSpecDB/msdb.nam.Z";
119     $taxonomy_file_url[0] = "ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz";
120     $version_regex = 'MSDB\s+Release\s+(\d+) ';
121 #
122 # NCBIInr_from_NCBI
123 #
124 } elsif ($ARGV[0] eq "NCBIInr_from_NCBI") {
125     $db_name = "NCBIInr";
126     $local_incoming_directory = "$MASCOT/sequence/NCBIInr/incoming";
127 #
128 #
129 #
130 #
131 #
132 #
133 #
134 #
135 #
136 #
137 #
138 #
139 #
140 #
141 #
142 #
143 #
144 #
145 #
146 #
147 #
148 #
149 #
150 #
151 #
152 #
153 #
154 #
155 #
156 #
157 #
158 #
159 #
160 #
161 #
162 #
163 #
164 #
165 #
166 #
167 #
168 #
169 #
170 #
171 #
172 #
173 #
174 #
175 #
176 #
177 #
178 #
179 #
180 #
181 #
182 #
183 #
184 #
185 #
186 #
187 #
188 #
189 #
190 #
191 #
192 #
193 #
194 #
195 #
196 #
197 #
198 #
199 #
200 #
201 #
202 #
203 #
204 #
205 #
206 #
207 #
208 #
209 #
210 #
211 #
212 #
213 #
214 #
215 #
216 #
217 #
218 #
219 #
220 #
221 #
222 #
223 #
224 #
225 #
226 #
227 #
228 #
229 #
230 #
231 #
232 #
233 #
234 #
235 #
236 #
237 #
238 #
239 #
240 #
241 #
242 #
243 #
244 #
245 #
246 #
247 #
248 #
249 #
250 #
251 #
252 #
253 #
254 #
255 #
256 #
257 #
258 #
259 #
260 #
261 #
262 #
263 #
264 #
265 #
266 #
267 #
268 #
269 #
270 #
271 #
272 #
273 #
274 #
275 #
276 #
277 #
278 #
279 #
280 #
281 #
282 #
283 #
284 #
285 #
286 #
287 #
288 #
289 #
290 #
291 #
292 #
293 #
294 #
295 #
296 #
297 #
298 #
299 #
300 #
301 #
302 #
303 #
304 #
305 #
306 #
307 #
308 #
309 #
310 #
311 #
312 #
313 #
314 #
315 #
316 #
317 #
318 #
319 #
320 #
321 #
322 #
323 #
324 #
325 #
326 #
327 #
328 #
329 #
330 #
331 #
332 #
333 #
334 #
335 #
336 #
337 #
338 #
339 #
340 #
341 #
342 #
343 #
344 #
345 #
346 #
347 #
348 #
349 #
350 #
351 #
352 #
353 #
354 #
355 #
356 #
357 #
358 #
359 #
360 #
361 #
362 #
363 #
364 #
365 #
366 #
367 #
368 #
369 #
370 #
371 #
372 #
373 #
374 #
375 #
376 #
377 #
378 #
379 #
380 #
381 #
382 #
383 #
384 #
385 #
386 #
387 #
388 #
389 #
390 #
391 #
392 #
393 #
394 #
395 #
396 #
397 #
398 #
399 #
400 #
401 #
402 #
403 #
404 #
405 #
406 #
407 #
408 #
409 #
410 #
411 #
412 #
413 #
414 #
415 #
416 #
417 #
418 #
419 #
420 #
421 #
422 #
423 #
424 #
425 #
426 #
427 #
428 #
429 #
430 #
431 #
432 #
433 #
434 #
435 #
436 #
437 #
438 #
439 #
440 #
441 #
442 #
443 #
444 #
445 #
446 #
447 #
448 #
449 #
450 #
451 #
452 #
453 #
454 #
455 #
456 #
457 #
458 #
459 #
460 #
461 #
462 #
463 #
464 #
465 #
466 #
467 #
468 #
469 #
470 #
471 #
472 #
473 #
474 #
475 #
476 #
477 #
478 #
479 #
480 #
481 #
482 #
483 #
484 #
485 #
486 #
487 #
488 #
489 #
490 #
491 #
492 #
493 #
494 #
495 #
496 #
497 #
498 #
499 #
500 #
501 #
502 #
503 #
504 #
505 #
506 #
507 #
508 #
509 #
510 #
511 #
512 #
513 #
514 #
515 #
516 #
517 #
518 #
519 #
520 #
521 #
522 #
523 #
524 #
525 #
526 #
527 #
528 #
529 #
530 #
531 #
532 #
533 #
534 #
535 #
536 #
537 #
538 #
539 #
540 #
541 #
542 #
543 #
544 #
545 #
546 #
547 #
548 #
549 #
550 #
551 #
552 #
553 #
554 #
555 #
556 #
557 #
558 #
559 #
560 #
561 #
562 #
563 #
564 #
565 #
566 #
567 #
568 #
569 #
570 #
571 #
572 #
573 #
574 #
575 #
576 #
577 #
578 #
579 #
580 #
581 #
582 #
583 #
584 #
585 #
586 #
587 #
588 #
589 #
590 #
591 #
592 #
593 #
594 #
595 #
596 #
597 #
598 #
599 #
600 #
601 #
602 #
603 #
604 #
605 #
606 #
607 #
608 #
609 #
610 #
611 #
612 #
613 #
614 #
615 #
616 #
617 #
618 #
619 #
620 #
621 #
622 #
623 #
624 #
625 #
626 #
627 #
628 #
629 #
630 #
631 #
632 #
633 #
634 #
635 #
636 #
637 #
638 #
639 #
640 #
641 #
642 #
643 #
644 #
645 #
646 #
647 #
648 #
649 #
650 #
651 #
652 #
653 #
654 #
655 #
656 #
657 #
658 #
659 #
660 #
661 #
662 #
663 #
664 #
665 #
666 #
667 #
668 #
669 #
670 #
671 #
672 #
673 #
674 #
675 #
676 #
677 #
678 #
679 #
680 #
681 #
682 #
683 #
684 #
685 #
686 #
687 #
688 #
689 #
690 #
691 #
692 #
693 #
694 #
695 #
696 #
697 #
698 #
699 #
700 #
701 #
702 #
703 #
704 #
705 #
706 #
707 #
708 #
709 #
710 #
711 #
712 #
713 #
714 #
715 #
716 #
717 #
718 #
719 #
720 #
721 #
722 #
723 #
724 #
725 #
726 #
727 #
728 #
729 #
730 #
731 #
732 #
733 #
734 #
735 #
736 #
737 #
738 #
739 #
740 #
741 #
742 #
743 #
744 #
745 #
746 #
747 #
748 #
749 #
750 #
751 #
752 #
753 #
754 #
755 #
756 #
757 #
758 #
759 #
760 #
761 #
762 #
763 #
764 #
765 #
766 #
767 #
768 #
769 #
770 #
771 #
772 #
773 #
774 #
775 #
776 #
777 #
778 #
779 #
780 #
781 #
782 #
783 #
784 #
785 #
786 #
787 #
788 #
789 #
790 #
791 #
792 #
793 #
794 #
795 #
796 #
797 #
798 #
799 #
800 #
801 #
802 #
803 #
804 #
805 #
806 #
807 #
808 #
809 #
810 #
811 #
812 #
813 #
814 #
815 #
816 #
817 #
818 #
819 #
820 #
821 #
822 #
823 #
824 #
825 #
826 #
827 #
828 #
829 #
830 #
831 #
832 #
833 #
834 #
835 #
836 #
837 #
838 #
839 #
840 #
841 #
842 #
843 #
844 #
845 #
846 #
847 #
848 #
849 #
850 #
851 #
852 #
853 #
854 #
855 #
856 #
857 #
858 #
859 #
860 #
861 #
862 #
863 #
864 #
865 #
866 #
867 #
868 #
869 #
870 #
871 #
872 #
873 #
874 #
875 #
876 #
877 #
878 #
879 #
880 #
881 #
882 #
883 #
884 #
885 #
886 #
887 #
888 #
889 #
890 #
891 #
892 #
893 #
894 #
895 #
896 #
897 #
898 #
899 #
900 #
901 #
902 #
903 #
904 #
905 #
906 #
907 #
908 #
909 #
910 #
911 #
912 #
913 #
914 #
915 #
916 #
917 #
918 #
919 #
920 #
921 #
922 #
923 #
924 #
925 #
926 #
927 #
928 #
929 #
930 #
931 #
932 #
933 #
934 #
935 #
936 #
937 #
938 #
939 #
940 #
941 #
942 #
943 #
944 #
945 #
946 #
947 #
948 #
949 #
950 #
951 #
952 #
953 #
954 #
955 #
956 #
957 #
958 #
959 #
960 #
961 #
962 #
963 #
964 #
965 #
966 #
967 #
968 #
969 #
970 #
971 #
972 #
973 #
974 #
975 #
976 #
977 #
978 #
979 #
980 #
981 #
982 #
983 #
984 #
985 #
986 #
987 #
988 #
989 #
990 #
991 #
992 #
993 #
994 #
995 #
996 #
997 #
998 #
999 #
1000 #
1001 #
1002 #
1003 #
1004 #
1005 #
1006 #
1007 #
1008 #
1009 #
1010 #
1011 #
1012 #
1013 #
1014 #
1015 #
1016 #
1017 #
1018 #
1019 #
1020 #
1021 #
1022 #
1023 #
1024 #
1025 #
1026 #
1027 #
1028 #
1029 #
1030 #
1031 #
1032 #
1033 #
1034 #
1035 #
1036 #
1037 #
1038 #
1039 #
1040 #
1041 #
1042 #
1043 #
1044 #
1045 #
1046 #
1047 #
1048 #
1049 #
1050 #
1051 #
1052 #
1053 #
1054 #
1055 #
1056 #
1057 #
1058 #
1059 #
1060 #
1061 #
1062 #
1063 #
1064 #
1065 #
1066 #
1067 #
1068 #
1069 #
1070 #
1071 #
1072 #
1073 #
1074 #
1075 #
1076 #
1077 #
1078 #
1079 #
1080 #
1081 #
1082 #
1083 #
1084 #
1085 #
1086 #
1087 #
1088 #
1089 #
1090 #
1091 #
1092 #
1093 #
1094 #
1095 #
1096 #
1097 #
1098 #
1099 #
1100 #
1101 #
1102 #
1103 #
1104 #
1105 #
1106 #
1107 #
1108 #
1109 #
1110 #
1111 #
1112 #
1113 #
1114 #
1115 #
1116 #
1117 #
1118 #
1119 #
1120 #
1121 #
1122 #
1123 #
1124 #
1125 #
1126 #
1127 #
1128 #
1129 #
1130 #
1131 #
1132 #
1133 #
1134 #
1135 #
1136 #
1137 #
1138 #
1139 #
1140 #
1141 #
1142 #
1143 #
1144 #
1145 #
1146 #
1147 #
1148 #
1149 #
1150 #
1151 #
1152 #
1153 #
1154 #
1155 #
1156 #
1157 #
1158 #
1159 #
1160 #
1161 #
1162 #
1163 #
1164 #
1165 #
1166 #
1167 #
1168 #
1169 #
1170 #
1171 #
1172 #
1173 #
1174 #
1175 #
1176 #
1177 #
1178 #
1179 #
1180 #
1181 #
1182 #
1183 #
1184 #
1185 #
1186 #
1187 #
1188 #
1189 #
1190 #
1191 #
1192 #
1193 #
1194 #
1195 #
1196 #
1197 #
1198 #
1199 #
1200 #
1201 #
1202 #
1203 #
1204 #
1205 #
1206 #
1207 #
1208 #
1209 #
1210 #
1211 #
1212 #
1213 #
1214 #
1215 #
1216 #
1217 #
1218 #
1219 #
1220 #
1221 #
1222 #
1223 #
1224 #
1225 #
1226 #
1227 #
1228 #
1229 #
1230 #
1231 #
1232 #
1233 #
1234 #
1235 #
1236 #
1237 #
1238 #
1239 #
1240 #
1241 #
1242 #
1243 #
1244 #
1245 #
1246 #
1247 #
1248 #
1249 #
1250 #
1251 #
1252 #
1253 #
1254 #
1255 #
1256 #
1257 #
1258 #
1259 #
1260 #
1261 #
1262 #
1263 #
1264 #
1265 #
1266 #
1267 #
1268 #
1269 #
1270 #
1271 #
1272 #
1273 #
1274 #
1275 #
1276 #
1277 #
1278 #
1279 #
1280 #
1281 #
1282 #
1283 #
1284 #
1285 #
1286 #
1287 #
1288 #
1289 #
1290 #
1291 #
1292 #
1293 #
1294 #
1295 #
1296 #
1297 #
1298 #
1299 #
1300 #
1301 #
1302 #
1303 #
1304 #
1305 #
1306 #
1307 #
1308 #
1309 #
1310 #
1311 #
1312 #
1313 #
1314 #
1315 #
1316 #
1317 #
1318 #
1319 #
1320 #
1321 #
1322 #
1323 #
1324 #
1325 #
1326 #
1327 #
1328 #
1329 #
1330 #
1331 #
1332 #
1333 #
1334 #
1335 #
1336 #
1337 #
1338 #
1339 #
1340 #
1341 #
1342 #
1343 #
1344 #
1345 #
1346 #
1347 #
1348 #
1349 #
1350 #
1351 #
1352 #
1353 #
1354 #
1355 #
1356 #
1357 #
1358 #
1359 #
1360 #
1361 #
1362 #
1363 #
1364 #
1365 #
1366 #
1367 #
1368 #
1369 #
1370 #
1371 #
1372 #
1373 #
1374 #
1375 #
1376 #
1377 #
1378 #
1379 #
1380 #
1381 #
1382 #
1383 #
1384 #
1385 #
1386 #
1387 #
1388 #
1389 #
1390 #
1391 #
1392 #
1393 #
1394 #
1395 #
1396 #
1397 #
1398 #
1399 #
1400 #
1401 #
1402 #
1403 #
1404 #
1405 #
1406 #
1407 #
1408 #
1409 #
1410 #
1411 #
1412 #
1413 #
1414 #
1415 #
1416 #
1417 #
1418 #
1419 #
1420 #
1421 #
1422 #
1423 #
1424 #
1425 #
1426 #
1427 #
1428 #
1429 #
1430 #
1431 #
1432 #
1433 #
1434 #
1435 #
1436 #
1437 #
1438 #
1439 #
1440 #
1441 #
1442 #
1443 #
1444 #
1445 #
1446 #
1447 #
1448 #
1449 #
1450 #
1451 #
1452 #
1453 #
1454 #
1455 #
1456 #
1457 #
1458 #
1459 #
1460 #
1461 #
1462 #
1463 #
1464 #
1465 #
1466 #
1467 #
1468 #
1469 #
1470 #
1471 #
1472 #
1473 #
1474 #
1475 #
1476 #
1477 #
1478 #
1479 #
1480 #
1481 #
1482 #
1483 #
1484 #
1485 #
1486 #
1487 #
1488 #
1489 #
1490 #
1491 #
1492 #
1493 #
1494 #
1495 #
1496 #
1497 #
1498 #
1499 #
1500 #
1501 #
1502 #
1503 #
1504 #
1505 #
1506 #
1507 #
1508 #
1509 #
1510 #
1511 #
1512 #
1513 #
1514 #
1515 #
1516 #
1517 #
1518 #
1519 #
1520 #
1521 #
1522 #
1523 #
1524 #
1525 #
1526 #
1527 #
1528 #
1529 #
1530 #
1531 #
1532 #
1533 #
1534 #
1535 #
1536 #
1537 #
1538 #
1539 #
1540 #
1541 #
1542 #
1543 #
1544 #
1545 #
1546 #
1547 #
1548 #
1549 #
1550 #
1551 #
1552 #
1553 #
1554 #
1555 #
1556 #
1557 #
1558 #
1559 #
1560 #
1561 #
1562 #
1563 #
1564 #
1565 #
1566 #
1567 #
1568 #
1569 #
1570 #
1571 #
1572 #
1573 #
1574 #
1575 #
1576 #
1577 #
1578 #
1579 #
1580 #
1581 #
1582 #
1583 #
1584 #
1585 #
1586 #
1587 #
1588 #
1589 #
1590 #
1591 #
1592 #
1593 #
1594 #
1595 #
1596 #
1597 #
1598 #
1599 #
1600 #
1601 #
1602 #
1603 #
1604 #
1605 #
1606 #
1607 #
1608 #
1609 #
1610 #
1611 #
1612 #
1613 #
1614 #
1615 #
1616 #
1617 #
1618 #
1619 #
1620 #
1621 #
1622 #
1623 #
1624 #
1625 #
1626 #
1627 #
1628 #
1629 #
1630 #
1631 #
1632 #
1633 #
1634 #
1635 #
1636 #
1637 #
1638 #
1639 #
1640 #
1641 #
1642 #
1643 #
1644 #
1645 #
1646 #
1647 #
1648 #
1649 #
1650 #
1651 #
1652 #
1653 #
1654 #
1655 #
1656 #
1657 #
1658 #
1659 #
1660 #
1661 #
1662 #
1663 #
1664 #
1665 #
1666 #
1667 #
1668 #
1669 #
1670 #
1671 #
1672 #
1673 #
1674 #
1675 #
1676 #
1677 #
1678 #
1679 #
1680 #
1681 #
1682 #
1683 #
1684 #
1685 #
1686 #
1687 #
1688 #
1689 #
1690 #
1691 #
1692 #
1693 #
1694 #
1695 #
1696 #
1697 #
1698 #
1699 #
1700 #
1701 #
1702 #
1703 #
1704 #
1705 #
1706 #
1707 #
1708 #
1709 #
1710 #
1711 #
1712 #
1713 #
1714 #
1715 #
1716 #
1717 #
1718 #
1719 #
1720 #
1721 #
1722 #
1723 #
1724 #
1725 #
1726 #
1727 #
1728 #
1729 #
1730 #
1731 #
1732 #
1733 #
1734 #
1735 #
1736 #
1737 #
1738 #
1739 #
1740 #
1741 #
1742 #
1743 #
1744 #
1745 #
1746 #
1747 #
1748 #
1749 #
1750 #
1751 #
1752 #
1753 #
1754 #
1755 #
1756 #
1757 #
1758 #
1759 #
1760 #
1761 #
1762 #
1763 #
1764 #
1765 #
1766 #
1767 #
1768 #
1769 #
1770 #
1771 #
1772 #
1773 #
1774 #
1775 #
1776 #
1777 #
1778 #
1779 #
1780 #
1781 #
1782 #
1783 #
1784 #
1785 #
1786 #
1787 #
1788 #
1789 #
1790 #
1791 #
1792 #
1793 #
1794 #
1795 #
1796 #
1797 #
1798 #
1799 #
1800 #
1801 #
1802 #
1803 #
1804 #
1805 #
1806 #
1807 #
1808 #
1809 #
1810 #
1811 #
1812 #
1813 #
1814 #
1815 #
1816 #
1817 #
1818 #
1819 #
1820 #
1821 #
1822 #
1823 #
1824 #
1825 #
1826 #
1827 #
1828 #
1829 #
1830 #
1831 #
1832 #
1833 #
1834 #
1835 #
1836 #
1837 #
1838 #
1839 #
1840 #
1841 #
1842 #
1843 #
1844 #
1845 #
1846 #
1847 #
1848 #
1849 #
1850 #
1851 #
1852 #
1853 #
1854 #
1855 #
1856 #
1857 #
1858 #
1859 #
1860 #
1861 #
1862 #
1863 #
1864 #
1865 #
1866 #
1867 #
1868 #
1869 #
1870 #
1871 #
1872 #
1873 #
1874 #
1875 #
1876 #
1877 #
1878 #
1879 #
1880 #
1881 #
1882 #
1883 #
1884 #
1885 #
1886 #
1887 #
1888 #
1889 #
1890 #
1891 #
1892 #
1893 #
1894 #
1895 #
1896 #
1897 #
1898 #
1899 #
1900 #
1901 #
1902 #
1903 #
1904 #
1905 #
1906 #
1907 #
1908 #
1909 #
1910 #
1911 #
1912 #
1913 #
1914 #
1915 #
1916 #
1917 #
1918 #
1919 #
1920 #
1921 #
1922 #
1923 #
1924 #
1925 #
1926 #
1927 #
1928 #
1929 #
1930 #
1931 #
1932 #
1933 #
1934 #
1935 #
1936 #
1937 #
1938 #
1939 #
1940 #
1941 #
1942 #
1943 #
1944 #
1945 #
1946 #
1947 #
1948 #
1949 #
1950 #
1951 #
1952 #
1953 #
1954 #
1955 #
1956 #
1957 #
1958 #
1959 #
1960 #
1961 #
1962 #
1963 #
1964 #
1965 #
1966 #
1967 #
1968 #
1969 #
1970 #
1971 #
1972 #
1973 #
1974 #
1975 #
1976 #
1977 #
1978 #
1979 #
1980 #
1981 #
1982 #
1983 #
1984 #
1985 #
1986 #
1987 #
1988 #
1989 #
1990 #
1991 #
1992 #
1993 #
1994 #
1995 #
1996 #
1997 #
1998 #
1999 #
2000 #
2001 #
2002 #
2003 #
2004 #
2005 #
2006 #
2007 #
2008 #
2009 #
2010 #
2011 #
2012 #
2013 #
2014 #
2015 #
2016 #
2017 #
2018 #
2019 #
2020 #
2021 #
2022 #
2023 #
2024 #
2025 #
2026 #
2027 #
2028 #
2029 #
2030 #
2031 #
2032 #
2033 #
2034 #
2035 #
2036 #
2037 #
2038 #
2039 #
2040 #
2041 #
2042 #
2043 #
2044 #
2045 #
2046 #
2047 #
2048 #
2049 #
2050 #
2051 #
2052 #
2053 #
2054 #
2055 #
2056 #
2057 #
2058 #
2059 #
2060 #
2061 #
2062 #
2063 #
2064 #
2065 #
2066 #
2067 #
2068 #
2069 #
2070 #
2071 #
2072 #
2073 #
2074 #
2075 #
2076 #
2077 #
2078 #
2079 #
2080 #
2081 #
2082 #
2083 #
2084 #
2085 #
2086 #
2087 #
2088 #
2089 #
2090 #
2091 #
2092 #
2093 #
2094 #
2095 #
2096 #
2097 #
2098 #
2099 #
2100 #
2101 #
2102 #
2103 #
2104 #
2105 #
2106 #
2107 #
2108 #
2109 #
2110 #
2111 #
2112 #
2113 #
2114 #
2115 #
2116 #
2117 #
2118 #
2119 #
2120 #
2121 #
2122 #
2123 #
2124 #
2125 #
2126 #
2127 #
2128 #
2129 #
2130 #
2131 #
2132 #
2133 #
2134 #
2135 #
2136 #
2137 #
2138 #
2139 #
2140 #
2141 #
2142 #
2143 #
2144 #
2145 #
2146 #
2147 #
2148 #
2149 #
2150 #
2151 #
2152 #
2153 #
2154 #
2155 #
2156 #
2157 #
2158 #
2159 #
2160 #
2161 #
2162 #
2163 #
2164 #
2165 #
2166 #
2167 #
2168 #
2169 #
2170 #
2171 #
2172 #
2173 #
2174 #
2175 #
2176 #
2177 #
2178 #
2179 #
2180 #
2181 #
2182 #
2183 #
2184 #
2185 #
2186 #
2187 #
2188 #
2189 #
2190 #
2191 #
2192 #
2193 #
2194 #
2195 #
2196 #
2197 #
2198 #
2199 #
2200 #
2201 #
2202 #
2203 #
2204 #
2205 #
2206 #
2207 #
2208 #
2209 #
2210 #
2211 #
2212 #
2213 #
2214 #
2215 #
2
```

Predefined configurations

- MSDB
- NCBItr
- EST_human
- EST_mouse
- EST_others
- Sprout
- Trembl
- IPI_human
- IPI_mouse



```
Command Prompt
C:\Inetpub\MASCOT\bin>db_update.pl EST_others_from_MGBI
```

ASMS 2003



All these databases are predefined in the script, so to update all the files, you only need to type something like this at a command or shell prompt.

Adding a new database to the script is quite straight forward, just use one of these as a template

Automate with Cron or Scheduled Tasks

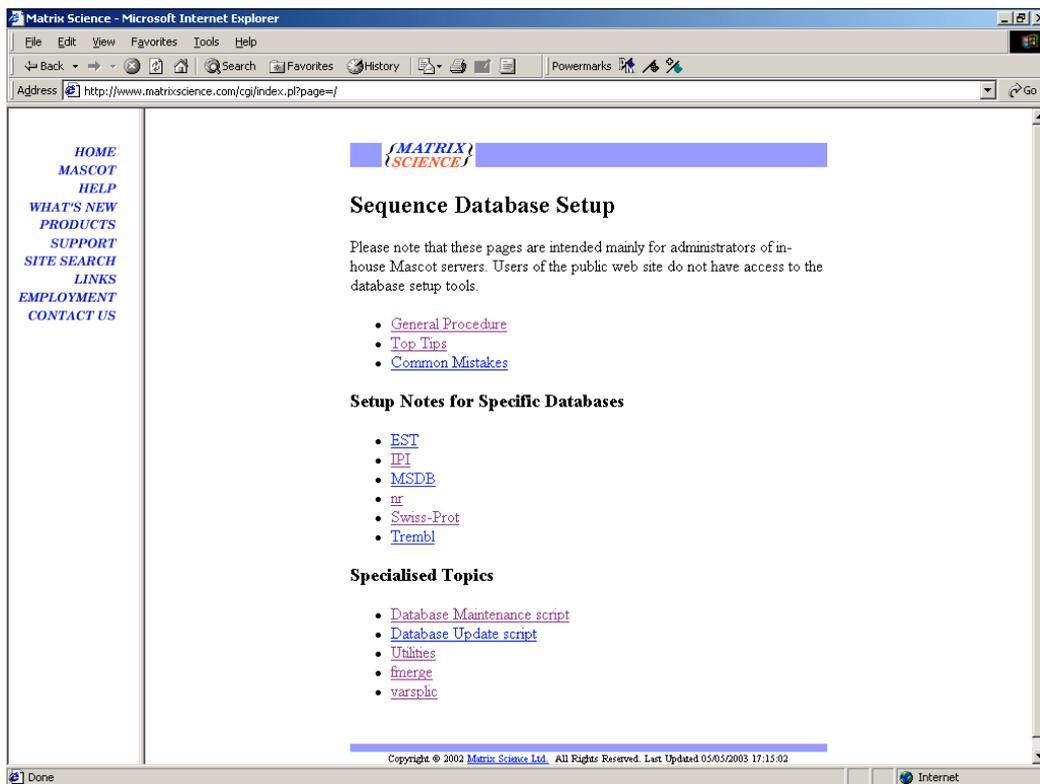
```
192.168.8.45 - PuTTY
$ crontab -l
# DO NOT EDIT THIS FILE - edit the master and reinstall.
# (crontab installed on Fri Aug  9 17:12:25 2002)
# (Cron version -- $FreeBSD: src/usr.sbin/cron/crontab/cron
p $)
00 4 * * * 0 /usr/local/www/bin/db_update
00 8 * * * 0 /usr/local/www/bin/db_update
00 10 * * * 0 /usr/local/www/bin/db_update
00 12 * * * 0 /usr/local/www/bin/db_update
00 16 * * * 0 /usr/local/www/bin/db_update
00 20 * * * 0 /usr/local/www/bin/db_update
$
```

The screenshot shows the 'Task Scheduler' window for a task named 'NCBIInr_from_NCBI'. The task is located at 'C:\WINNT\Tasks\NCBIInr_from_NCBI.job'. The 'Run' field contains 'perl db_update.pl NCBIInr_from_NCBI'. The 'Start in' field is 'C:\inetpub\wwwroot\bin\'. The 'Run as' field is 'MATRIX_SCIENCE\JohnC'. The 'Enabled' checkbox is checked. The 'OK', 'Cancel', and 'Apply' buttons are visible at the bottom.

ASMS 2003



Now, we are in a position to automate the whole process. Under Unix, we use Cron to execute the script on a pre-defined schedule
Under Windows, we use Scheduled Tasks



Finally, we have made an effort to provide detailed walk-throughs for configuring all the common databases. Please refer to the on-line help for greater detail on all the topics I've mentioned here