MARS and the Medical Text Indexer (MTI)

Test Bed

Initial Design:	June 12, 2001
Testing Period:	June 26, 2001 – July 12, 2001
Final Summary:	July 18, 2001

Table of Contents

Diagrams

1. Introduction

The purpose of this Test Bed is to develop a test environment where the Medical Text Indexer team can test out "real-time" processing and handling of OCCS (XML) formatted data files. The idea is to use the existing MARS automated data flow to OCCS as a platform to simulate the interaction with OCCS and allow the Medical Text Indexer team to finalize development of a system that will eventually interface directly with OCCS (see diagram below).

The system must be able to handle the test bed flow without hesitation, be able to run without human monitoring, provide near 100% uptime, and provide 100% accuracy integrating our recommendations into the existing data file without injecting inconsistencies in the data file.
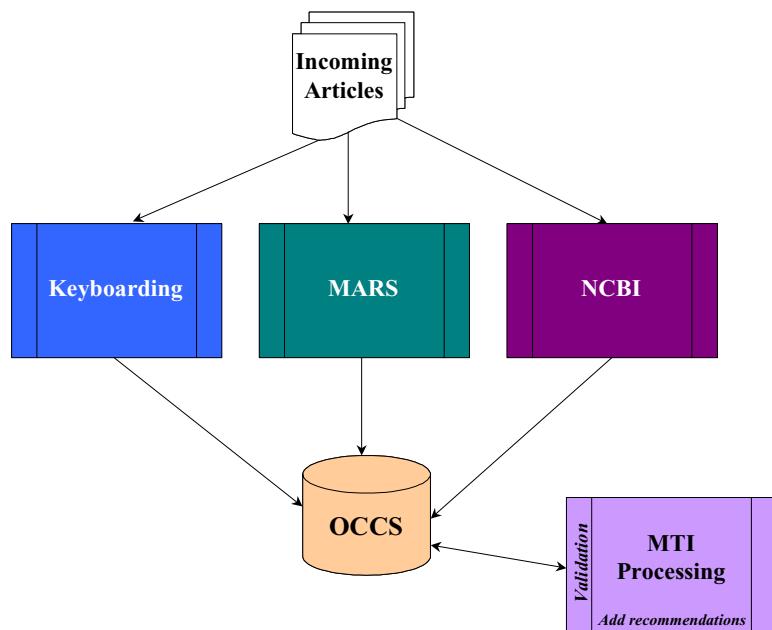
Diagram 1 - Long-term Goal of Test Bed

2. Summary of Test Runs/Findings

Overall, the MTI worked as planned. We were able to identify batch files sent from MARS, validate the input files, process all of the citations from the batch files, identify and integrate our recommendations, validate our reintegrated file, and transfer the batch files (now with MTI recommendations) back to MARS on average in less than five (5) minutes time.

Valuable information was also garnered as to "normal" operating procedures to expect in a production environment. We were able to identify typical batch sizes, beginning and ending of processing day, and that jobs are sporadic throughout the day versus all showing up at the end of day.

Thanks to the test bed, we were able to sail the MTI on a "shake-down cruise" identifying and fixing any problems that arose. During this time period we identified and corrected three problems with the MTI program:

1. Sporadic return of empty XML section tags

2. Wrong DTD used (nlm.dtd vs nlminputarticle.dtd)

3. A problem when the list of PubMed Related Citations returned containing no useable MeSH Headings. In a single case, all of the recommended citations from the Related Citations module were pre-Medline and contained no MeSH Headings from which to make suggestions from.

Two concerns have emerged as a result of our testing:

1. We currently have a single interface into the PubMed Related Citations API that is completely outside of our control. When this system is down, we only have the ability to send email to our point of contact at NCBI where the problem could be one of two things: 1) a problem with the API we are using, or 2) a problem with the back-end database. If this is a problem with the back-end database, our point of contact then contacts their point of contact in the database group and then information flows backwards along this chain which is a slow processing involving up to a days time.

2. What are we going to do in the event of either MetaMap or PubMed Related Citations becoming unavailable? Do we provide no recommendations and pass along the batch files? Do we provide recommendations with possibly less precision using the system that is currently available? Do we hold up processing until everything is fixed?

3. Background on the Medical Text Indexer
3.1. What it does

The Medical Text Indexer (MTI) is designed to process MEDLINE citations via several pathways (see diagram below), and produce a single ordered list of recommended indexing terms for the citation.
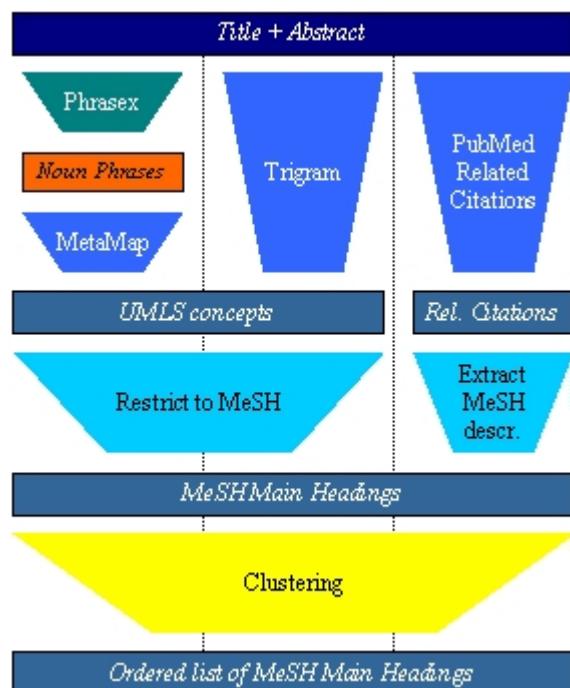

Diagram 2 – Medical Test Indexer

3.2. Performance

The current version of the MTI can handle approximately 200 average sized citations per hour during normal business hours and approximately 300 average sized citations per hour outside of normal business hours. An average size citation is defined as 1,900 bytes in length based on averages from citations reviewed for MEDLINE years 1997 - 1999. Non-business hours are defined as all day on Saturday and Sunday, and 10:00 PM to 6:00 AM during the weekdays. This is a limitation from the PubMed Related Citations pathway due to a throttle limiting the number of requests from a user during peak times, and will have to be looked at in the future.

4. Test Bed
4.1. Introduction

The creation of the test bed should require minimal work for all participants (see section *"4. Anticipated Work to Create Test Bed"* for details) and, most importantly, have minimal impact on the MARS production system. The test bed will involve the following:

1. Transmittal of final validated XML based file from MARS to MTI at approximately the same time as they are transmitted to OCCS.

2. Automatic detection and validation of the MARS transmitted file by the MTI.

3. Processing and final recommendations made by the MTI and reintegrated into the MARS file.

4. MTI Validation prior to transmittal back to MARS.

5. MTI transmittal of the file back to MARS.

6. MARS validation of the file from MTI - offline from production system. Batch processed with email every couple of days – more frequently at beginning tapering off towards the end of the experiment.
   a. Verification of receipt
   b. Validation of file format to ensure that no inconsistencies were added to the file.

7. Transmittal of MARS validation report back to MTI denoting any deficiencies found.
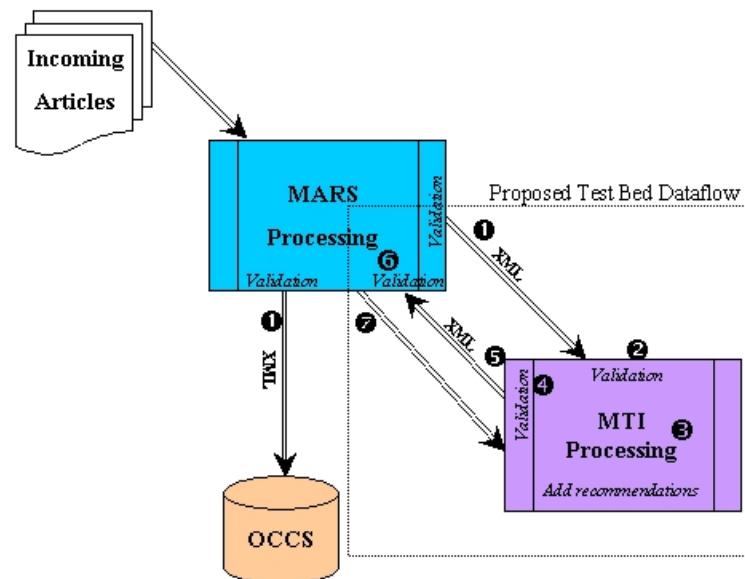
Diagram 3 - Proposed Test Bed

4.2. Timeline

Please note that all times are flexible.

Agreement on protocols	Beginning of June	
Initial Testing of Communications Interface	Duration: 1 – 5 days	Start: 2nd week of June
Live Operations	Duration: 1 month	Start: 3rd week of June
Status Meeting(s)	Duration: 15 – 30 minutes	As necessary during live operations time period.
Summary Meeting	Duration: 1 hour	End of July

Agreement on protocols

- Agree on how we are going to transfer files back and forth.

    o **FTP**

    o **Username:**

    o **Password:**

    o **FTP Address:**

- Agree on protocols – naming of files, where to send the information (email addresses, ftp addresses, etc.), and decide on validation requirements, etc.

Initial Testing of Communications Interface [1 – 5 days]:

- Start with the basic transmittal and acknowledgement of files via transmission mechanism with email acknowledgements – simple confirmation of ability to transmit and receive in both directions.

- Test transmission and processing of a data file to validate full cycle operation flow – repeat as necessary based on debugging and shakedown needs.

Live Operations [2 weeks]:

- Live testing of the MTI system with the MARS system. This is the actual automatic transmission of data to the MTI by the MARS system, receipt validation and processing by the MTI, validation and retransmittal by the MTI, and receipt and validation by the MARS system.

- A validation report will then be generated for the MTI processed files in batch offline from the main MARS production system. This report listing any deficiencies found in the MTI processed file will than be emailed to the MTI point-of-contact.

Status Meeting(s) [15 – 30 minutes as needed]:

- Discussion of any problems encountered by either MTI or MARS staff. Typically this will be handled by the appointed point-of-contact on an ad-hoc basis.

Summary Meeting [1 hour]:

- Wrap up with discussions on how testing went, suggestions for improvement, etc.

5. Anticipated Work to Create Test Bed
5.1. MARS Side

The anticipated work on the MARS side is expected to be light as we don't want to unduly burden the MARS team with work for us. It is anticipated that following will be needed:

- A developer to add or modify the existing automated system for ftping files to OCCS to also ftp our system at the same time.

- A developer to develop a validation program (hopefully a slight modification of the existing program) that will validate our files that are sent back to ensure that we did not insert any errors into the file. The program will be ran periodically offline from the main MARS production system and results will then be emailed back to MTI.

- A point-of-contact person to interface between the two teams to ensure proper communication.

5.2. Medical Text Indexer Side

The anticipated work on the Medical Text Indexer side is also expected to be light. It is anticipated that following will be needed:

- A developer to modify the existing automated system for recognizing ftp files and performing some operation on the file.

- A developer to develop an automated validation that will validate the incoming file as well as validate our file that is sent back to ensure that we did not insert any errors. If possible, get a copy of the MARS validation program and start there.

- A developer to modify existing system to handle queued processing of input data files with the ability to handle multi-tiered priorities for the data files to be processed.

- A person to monitor activity and ensure smooth processing.

- A point-of-contact person to interface between the two teams to ensure proper communication.

6. Proposed XML Tags

The list below details the full extent of XML tags we are proposing to insert into the MARS data file.

```
<IND_Recommendations>
    <MTI_Heading_List>
        <MTI_Heading>
            <MTI_Concept> </MTI_Concept>
            <MTI_Location> </MTI_Location>
        </MTI_Heading>
    </MTI_Heading_List>

    <MTI_SupplementaryConcept_List>
        <MTI_SupplementaryConcept>
            <MTI_Concept> </MTI_Concept>
            <MTI_Location> </MTI_Location>
        </MTI_SupplementaryConcept >
    </MTI_SupplementaryConcept_List>

    <MTI_CheckTag_List>
        <MTI_CheckTag>
            <MTI_Concept> </MTI_Concept>
            <MTI_Location> </MTI_Location>
        </MTI_CheckTag>
    </MTI_CheckTag_List>

    <MTI_SubHeading_List>
        <MTI_SubHeading>
            <MTI_Concept> </MTI_Concept>
            <MTI_Location> </MTI_Location>
        </MTI_SubHeading>
    </MTI_SubHeading_List>
</IND_Recommendations>
```

7. Proposed XML DTD Changes

The two file listings below highlight the proposed changes to the existing NLM Medline DTDs to accommodate the Medical Text Indexer fields. We have inserted changes into the "nlminputarticle.dtd" and "nlmcommon.dtd" files as shown by the **bold** font in the two listings below. **Special Note:** *Make sure to use the nlminputarticle.dtd versus the nlm.dtd.*

```
<!--    This is the Current DTD which NLM has written for
        Internal  Use.
        (December 11, 2000)

        If you are a Input User, use the information
        from the InputArticle Set.


        The SAMPLE FILE is located in
     "http://www.nlm.nih.gov/databases/dtd/nlminputarticle_sample.xml"
        -->
<!-- ================================================================ -->
<!ENTITY % ArticleTitle.Ref "ArticleTitle?">
<!ENTITY % ISSN.Ref "ISSN?">
<!ENTITY % PubDate.Ref "PubDate?">
<!-- ================================================================ -->
<!-- ================================================================ -->
<!-- Reference to Where the NLMCOMMON DTD is located  -->
<!ENTITY % NlmCommon PUBLIC "-//NLM//DTD Common, 9th
      May 2000//EN" "http://nls9.nlm.nih.gov/DTD/nlmcommon.dtd" >
%NlmCommon;
<!-- ================================================================ -->
<!ELEMENT InputArticleSet (InputArticle+)>
<!-- ================================================================ -->
<!-- This is the top level element for InputArticle -->
<!ELEMENT InputArticle (Article, MeshHeadingList?, IND_Recommendations)>
<!ATTLIST InputArticle
      AnonymousAuthorYN (Y | N) "N"
>
<!ATTLIST JournalIssue
      MRI CDATA #REQUIRED
>
<!-- End of InputArticle group -->
```

```
<!-- NLMCommon DTD

     This is the current DTD for data elements that are shared
     among various applications at the NLM.
     Comments and suggestions are welcome.
     December 11, 2000
-->
<!--    NLMCommon.dtd

        Document Type Definition for the PubMed Article DTD
        $Id$

        This version includes the addition of NLMUniqueID to the DTD.
-->
<!-- Personal and Author names -->
<!ENTITY % personal.name "(LastName, (FirstName, MiddleName?)?,
                          Initials?, Suffix?)">
<!ENTITY % author.name "(%personal.name; | CollectiveName)">
<!ELEMENT FirstName (#PCDATA)>
<!ELEMENT MiddleName (#PCDATA)>
<!ELEMENT LastName (#PCDATA)>
<!ELEMENT Initials (#PCDATA)>
<!ELEMENT Suffix (#PCDATA)>
<!ELEMENT CollectiveName (#PCDATA)>
<!-- Dates -->
<!ENTITY % normal.date "(Year,Month,Day,(Hour,(Minute,Second?)?)?)">
<!ENTITY % pub.date "((Year, ((Month, Day?) | Season)?) | MedlineDate)">
<!ELEMENT Year (#PCDATA)>
<!ELEMENT Month (#PCDATA)>
<!ELEMENT Day (#PCDATA)>
<!ELEMENT Season (#PCDATA)>
<!ELEMENT MedlineDate (#PCDATA)>
<!ELEMENT Hour (#PCDATA)>
<!ELEMENT Minute (#PCDATA)>
<!ELEMENT Second (#PCDATA)>
<!ENTITY % data.template "#PCDATA">
<!-- ================================================================= -->
<!-- ================================================================= -->
<!-- This is the top level element for NCBIArticle -->
<!ELEMENT NCBIArticle (PMID, Article, MedlineJournalInfo?)>
<!-- ================================================================= -->
<!-- This is the top level element for Article -->
<!ELEMENT Article ((Journal | Book),
                   %ArticleTitle.Ref;,
                   Pagination,
                   Abstract?,
                   Affiliation?,
                   AuthorList?,
                   Language+,
                   DataBankList?,
                   GrantList?,
                   PublicationTypeList,
                   VernacularTitle?,
                   DateOfElectronicPublication?)>
<!ELEMENT DataBankList (DataBank+)>
<!ELEMENT DataBank (DataBankName, AccessionNumberList?)>
<!ELEMENT DataBankName (#PCDATA)>
<!ELEMENT AccessionNumberList (AccessionNumber+)>
<!ELEMENT AccessionNumber (#PCDATA)>
<!ATTLIST DataBankList
        CompleteYN (Y | N) "Y"
>
```

```
<!ELEMENT GrantList (Grant+)>
<!ELEMENT Grant (GrantID, Acronym?, Agency?)>
<!ELEMENT GrantID (#PCDATA)>
<!ELEMENT Acronym (#PCDATA)>
<!ELEMENT Agency (#PCDATA)>
<!ATTLIST GrantList
        CompleteYN (Y | N) "Y"
>
<!ELEMENT Journal (%ISSN.Ref;,
                   JournalIssue,
                   Coden?,
                   Title?,
                   ISOAbbreviation?)>
<!ELEMENT ISSN (#PCDATA)>
<!ELEMENT JournalIssue (Volume?, Issue?, %PubDate.Ref;)>
<!ELEMENT Volume (#PCDATA)>
<!ELEMENT Issue (#PCDATA)>
<!ELEMENT PubDate (%pub.date;)>
<!ELEMENT Coden (#PCDATA)>
<!ELEMENT Title (#PCDATA)>
<!ELEMENT ISOAbbreviation (#PCDATA)>
<!ELEMENT DateOfElectronicPublication (#PCDATA)>
<!ELEMENT MedlineJournalInfo (Country,
                              MedlineTA,
                              MedlineCode,
                              NlmUniqueID?)>
<!ELEMENT Country (#PCDATA)>
<!ELEMENT MedlineTA (#PCDATA)>
<!ELEMENT MedlineCode (#PCDATA)>
<!-- Sometime in the future, MedlineCode will change to
     NLMUniqueID   -->
<!ELEMENT Book (%PubDate.Ref;,
                Publisher,
                Title,
                AuthorList?,
                CollectionTitle?,
                Volume?)>
<!ELEMENT Publisher (#PCDATA)>
<!ELEMENT ArticleTitle (#PCDATA)>
<!ELEMENT CollectionTitle (#PCDATA)>
<!ELEMENT VernacularTitle (#PCDATA)>
<!ELEMENT PublicationTypeList (PublicationType+)>
<!ELEMENT PublicationType (#PCDATA)>
<!ELEMENT Language (#PCDATA)>
<!ELEMENT AuthorList (Author+)>
<!ELEMENT Author ((%author.name;), Affiliation?)>
<!ELEMENT Affiliation (#PCDATA)>
<!ATTLIST AuthorList
        CompleteYN (Y | N) "Y"
>
<!ELEMENT Abstract (AbstractText, CopyrightInformation?)>
<!ELEMENT AbstractText (%data.template;)>
<!ELEMENT CopyrightInformation (#PCDATA)>
<!ELEMENT Pagination ((StartPage, EndPage?, MedlinePgn?) | MedlinePgn)>
<!ELEMENT StartPage (#PCDATA)>
<!ELEMENT EndPage (#PCDATA)>
<!ELEMENT MedlinePgn (#PCDATA)>
<!ELEMENT MeshHeadingList (MeshHeading+)>
<!ELEMENT MeshHeading (Descriptor, SubHeading*)>
<!ELEMENT Descriptor (#PCDATA)>
<!ATTLIST Descriptor
        MajorTopicYN (Y | N) "N"
>
```

```
<!ELEMENT SubHeading (#PCDATA)>
<!ATTLIST SubHeading
        MajorTopicYN (Y | N) "N"
>
<!ELEMENT PMID (#PCDATA)>
<!ELEMENT NlmUniqueID (#PCDATA)>
<!-- ================================================================= -->
<!-- ================================================================= -->
<!-- This is the top level element for Medical Text Indexer Recommendations -->
<!-- ================================================================= -->
<!-- This is the top level element for INDRecommendations -->
<!ELEMENT IND_Recommendations (MTI_Heading_List, MTI_SupplementaryConcept_List?,
                MTI_CheckTag_List?, MTI_SubHeading_List?)>

<!ELEMENT MTI_Heading_List (MTI_Heading+)>
<!ELEMENT MTI_SubHeading_List (MTI_SubHeading+)>
<!ELEMENT MTI_SupplementaryConcept_List (MTI_SupplementaryConcept+)>
<!ELEMENT MTI_CheckTag_List (MTI_CheckTag+)>

<!ELEMENT MTI_Heading (MTI_Concept, MTI_Location?)>
<!ELEMENT MTI_SubHeading (MTI_Concept, MTI_Location?)>
<!ELEMENT MTI_SupplementaryConcept (MTI_Concept, MTI_Location?)>
<!ELEMENT MTI_CheckTag (MTI_Concept, MTI_Location?)>

<!ELEMENT MTI_Concept (#PCDATA)>
<!ELEMENT MTI_Location (#PCDATA)>
```

8. Summary of Testing Results

Date	# Batch Jobs	# Items Processed	Total Proc. Time (in mins)	Avg. Proc. Time/ Item (in secs)	# Invalid Inputs	# MTI Errors	Start Time	Stop Time	Avg. # Items/Batch	Avg. Proc. Time/Batch (in mins)
6/29/2001	25	301	110.90	22.11	1	5	10:12:55	16:30:07	12.54	4.62
7/2/2001	33	405	128.25	19.00	2	2	7:58:17	16:38:17	13.06	4.14
7/3/2001	41	459	197.38	25.80	0	0	7:58:24	16:27:47	11.20	4.81
7/5/2001	34	515	163.32	19.03	0	3[*]	8:18:42	16:31:16	15.15	4.80
7/6/2001	32	481	207.72	25.91	0	0	8:37:46	16:27:09	15.03	6.49
7/9/2001	38	541	191.55	21.24	0	0	7:59:32	16:29:32	14.24	5.04
7/10/2001	26	363	119.12	19.69	0	1	7:55:48	15:50:19	13.96	4.58
7/11/2001	7	84	32.50	23.21	0	0	11:10:54	14:40:42	12.00	4.64
7/12/2001	43	721	219.05	18.23	0	0	6:08:54	15:51:55	16.77	5.09
Totals	279	3870	1,369.78	21.24	3	8				

Notes:

[*]**7/5/2001** - The 3 errors are actually related citations (relcit) problems where the single interface that we have was being used by someone else and we just had to restart the processing of the item and not really errors in the program.

7/6/2001 - Discovered that file_cnt program not counting articles with AnonymousAuthor values - modified file_cnt and reran on all jobs and updated the counts.

7/6/2001 - Long batch times may be due to large workload on nls9 - mmtx db loading and reconfiguring.

7/9/2001 - Problems with PubMed Related Citations - ran all with MetaMap only

7/10/2001 - Unforseen problem with empty return from PubMed Related Citations causing exit.

7/12/2001 - 16 Jobs all showed up at 0608 -- probably left over from the shortened 11th yesterday. NOTE: Heavy Workstation Load as well today.

Observations:

§ Jobs are typically run from just before 0800 in the morning to just after 1630 in the afternoon – but can come at any time. Anything out of this normal cycle is typically due to a processing problem on the MARS side but still must be accounted for in any production system.

§ Jobs are not run on weekends or holidays.

§ On average it takes us less than 5 minutes to process a batch file.

§ Jobs seem to come in randomly throughout the day.

§ Decision needs to be made in the event that either MetaMap or PubMed Related Citations is not available. Do we continue to run with the available services, do we pass in blank recommendations, or do we wait until everything is back up and running?

Detailed Results of Daily Batch Jobs between MARS and MTI

9. Detailed Testing Results

Date	Batch #	Item Count	Process Time (in secs.)	Process Time (in mins.)	Avg. Time Per Item	Notes
6/29/2001	1	14	354	5.90	25.29	MTI_CheckTag Error - Fixed - problem with empty section added incorrectly
	2	5	78	1.30	15.60	
	3	13	0	0.00	0.00	Invalid Input
	4	9	408	6.80	45.33	
	5	14	288	4.80	20.57	
	6	13	262	4.37	20.15	
	7	6	137	2.28	22.83	MTI_CheckTag Error - Fixed - problem with empty section added incorrectly
	8	15	312	5.20	20.80	
	9	5	136	2.27	27.20	
	10	6	146	2.43	24.33	
	11	12	304	5.07	25.33	
	12	7	174	2.90	24.86	
	13	7	148	2.47	21.14	
	14	8	194	3.23	24.25	
	15	18	357	5.95	19.83	
	16	29	637	10.62	21.97	AnonymousAuthorYN error - Fixed - problem of pointing to wrong DTD file
	17	13	295	4.92	22.69	
	18	10	278	4.63	27.80	
	19	9	321	5.35	35.67	
	20	34	789	13.15	23.21	
	21	9	109	1.82	12.11	
	22	22	445	7.42	20.23	
	23	11	192	3.20	17.45	MTI_CheckTag Error - Fixed - problem with empty section added incorrectly
	24	7	169	2.82	24.14	MTI_CheckTag Error - Fixed - problem with empty section added incorrectly
	25	5	121	2.02	24.20	
		301	6,654		22.11	
	Avg:	12.54	277	4.62		

Detailed Results of Daily Batch Jobs between MARS and MTI

Date	Batch #	Item Count	Process Time (in secs.)	Process Time (in mins.)	Avg. Time Per Item	Notes
7/2/2001	1	11	348	5.80	31.64	
	2	8	185	3.08	23.13	
	3	13	181	3.02	13.92	
	4	6	138	2.30	23.00	
	5	9	199	3.32	22.11	
	6	7	130	2.17	18.57	
	7	9	220	3.67	24.44	
	8	6	118	1.97	19.67	
	9	14	271	4.52	19.36	
	10	17	321	5.35	18.88	
	11	14	304	5.07	21.71	
	12	17	308	5.13	18.12	
	13	3	54	0.90	18.00	AnonymousAuthorYN error - Fixed - problem of pointing to wrong DTD file
	14	10	166	2.77	16.60	
	15	7	211	3.52	30.14	
	16	13	284	4.73	21.85	AnonymousAuthorYN error - Fixed - problem of pointing to wrong DTD file
	17	6	199	3.32	33.17	
	18	15	379	6.32	25.27	
	19	13	0	0.00		Invalid Input
	20	7	52	0.87	7.43	
	21	25	466	7.77	18.64	
	22	26	475	7.92	18.27	
	23	14	257	4.28	18.36	
	24	15	326	5.43	21.73	
	25	6	128	2.13	21.33	
	26	17	293	4.88	17.24	
	27	10	300	5.00	30.00	
	28	10	173	2.88	17.30	
	29	34	777	12.95	22.85	
	30	8	315	5.25	39.38	
	31	6	33	0.55	5.50	
	32	7	84	1.40	12.00	
	33	22	0	0.00		Invalid Input
		405	7,695		19.00	
	Avg:	13.06	248	4.14		

Detailed Results of Daily Batch Jobs between MARS and MTI

Date	Batch #	Item Count	Process Time (in secs.)	Process Time (in mins.)	Avg. Time Per Item	Notes
7/3/2001	1	15	325	5.42	21.67	
	2	5	147	2.45	29.40	
	3	14	195	3.25	13.93	
	4	3	553	9.22	184.33	
	5	7	117	1.95	16.71	
	6	10	212	3.53	21.20	
	7	6	82	1.37	13.67	
	8	9	226	3.77	25.11	
	9	7	249	4.15	35.57	
	10	6	331	5.52	55.17	
	11	11	107	1.78	9.73	
	12	11	282	4.70	25.64	
	13	8	236	3.93	29.50	
	14	19	478	7.97	25.16	
	15	12	256	4.27	21.33	
	16	21	604	10.07	28.76	
	17	9	214	3.57	23.78	
	18	6	151	2.52	25.17	
	19	10	134	2.23	13.40	
	20	12	408	6.80	34.00	
	21	7	134	2.23	19.14	
	22	14	417	6.95	29.79	
	23	18	445	7.42	24.72	
	24	16	404	6.73	25.25	
	25	14	284	4.73	20.29	
	26	8	347	5.78	43.38	
	27	21	320	5.33	15.24	
	28	18	426	7.10	23.67	
	29	10	264	4.40	26.40	
	30	5	106	1.77	21.20	
	31	4	167	2.78	41.75	
	32	10	272	4.53	27.20	
	33	20	357	5.95	17.85	
	34	20	451	7.52	22.55	
	35	15	384	6.40	25.60	
	36	10	435	7.25	43.50	
	37	9	213	3.55	23.67	
	38	5	240	4.00	48.00	
	39	19	541	9.02	28.47	
	40	9	224	3.73	24.89	
	41	6	105	1.75	17.50	
		459	11,843		25.80	
	Avg:	11.20	289	4.81		

Detailed Results of Daily Batch Jobs between MARS and MTI

Date	Batch #	Item Count	Process Time (in secs.)	Process Time (in mins.)	Avg. Time Per Item	Notes
7/5/2001	1	22	480	8.00	21.82	
	2	10	157	2.62	15.70	
	3	5	104	1.73	20.80	relcit problem - restarted and worked fine
	4	7	314	5.23	44.86	
	5	15	396	6.60	26.40	relcit problem - restarted and worked fine
	6	13	280	4.67	21.54	
	7	13	332	5.53	25.54	
	8	10	214	3.57	21.40	
	9	8	166	2.77	20.75	
	10	14	295	4.92	21.07	relcit problem - restarted and worked fine
	11	10	57	0.95	5.70	
	12	16	342	5.70	21.38	
	13	11	167	2.78	15.18	
	14	16	299	4.98	18.69	
	15	8	138	2.30	17.25	
	16	15	260	4.33	17.33	
	17	7	114	1.90	16.29	
	18	22	484	8.07	22.00	
	19	13	208	3.47	16.00	
	20	24	483	8.05	20.13	
	21	17	241	4.02	14.18	
	22	19	322	5.37	16.95	
	23	30	537	8.95	17.90	
	24	19	305	5.08	16.05	
	25	48	765	12.75	15.94	
	26	13	168	2.80	12.92	
	27	11	247	4.12	22.45	
	28	9	255	4.25	28.33	
	29	37	757	12.62	20.46	
	30	23	532	8.87	23.13	
	31	7	95	1.58	13.57	
	32	7	68	1.13	9.71	
	33	8	96	1.60	12.00	
	34	8	121	2.02	15.13	
		515	9,799		19.03	
	Avg:	15.15	288	4.80		

Detailed Results of Daily Batch Jobs between MARS and MTI

Date	Batch #	Item Count	Process Time (in secs.)	Process Time (in mins.)	Avg. Time Per Item	Notes
7/6/2001	1	10	207	3.45	20.70	
	2	16	411	6.85	25.69	
	3	15	285	4.75	19.00	
	4	12	238	3.97	19.83	
	5	11	223	3.72	20.27	
	6	21	477	7.95	22.71	
	7	11	288	4.80	26.18	
	8	11	356	5.93	32.36	
	9	10	148	2.47	14.80	
	10	11	242	4.03	22.00	
	11	18	433	7.22	24.06	
	12	12	315	5.25	26.25	
	13	14	320	5.33	22.86	
	14	8	193	3.22	24.13	
	15	71	1,129	18.82	15.90	
	16	6	110	1.83	18.33	
	17	17	275	4.58	16.18	
	18	13	132	2.20	10.15	
	19	18	3,194	53.23	177.44	
	20	25	391	6.52	15.64	
	21	25	450	7.50	18.00	
	22	13	303	5.05	23.31	
	23	22	416	6.93	18.91	
	24	8	138	2.30	17.25	
	25	10	435	7.25	43.50	
	26	1	49	0.82	49.00	
	27	26	442	7.37	17.00	
	28	18	502	8.37	27.89	
	29	9	79	1.32	8.78	
	30	9	97	1.62	10.78	
	31	8	135	2.25	16.88	
	32	2	50	0.83	25.00	
		481	12,463		25.91	
	Avg:	15.03	389	6.49		

Detailed Results of Daily Batch Jobs between MARS and MTI

Date	Batch #	Item Count	Process Time (in secs.)	Process Time (in mins.)	Avg. Time Per Item	Notes
7/9/2001	1	11	4939	82.32	449.00	Problems with Related Citations program returning results very slowly
	2	4	79	1.32	19.75	Switched over to MetaMap only from here on out for today
	3	9	209	3.48	23.22	
	4	14	212	3.53	15.14	
	5	12	135	2.25	11.25	
	6	20	241	4.02	12.05	
	7	16	370	6.17	23.13	
	8	14	135	2.25	9.64	
	9	8	151	2.52	18.88	
	10	32	111	1.85	3.47	
	11	23	201	3.35	8.74	
	12	18	243	4.05	13.50	
	13	5	42	0.70	8.40	
	14	42	221	3.68	5.26	
	15	6	140	2.33	23.33	
	16	13	247	4.12	19.00	
	17	23	429	7.15	18.65	
	18	25	226	3.77	9.04	
	19	5	84	1.40	16.80	
	20	9	126	2.10	14.00	
	21	14	242	4.03	17.29	
	22	10	73	1.22	7.30	
	23	28	143	2.38	5.11	
	24	10	218	3.63	21.80	
	25	15	200	3.33	13.33	
	26	16	256	4.27	16.00	
	27	14	114	1.90	8.14	
	28	19	323	5.38	17.00	
	29	6	53	0.88	8.83	
	30	14	212	3.53	15.14	
	31	15	143	2.38	9.53	
	32	3	30	0.50	10.00	
	33	18	250	4.17	13.89	
	34	11	111	1.85	10.09	
	35	4	37	0.62	9.25	
	36	8	197	3.28	24.63	
	37	21	224	3.73	10.67	
	38	6	126	2.10	21.00	
		541	11,493		21.24	
	Avg:	14.24	302	5.04		

Detailed Results of Daily Batch Jobs between MARS and MTI

Date	Batch #	Item Count	Process Time (in secs.)	Process Time (in mins.)	Avg. Time Per Item	Notes
7/10/2001	1	15	271	4.52	18.07	
	2	10	204	3.40	20.40	
	3	21	455	7.58	21.67	
	4	11	218	3.63	19.82	
	5	28	341	5.68	12.18	
	6	9	166	2.77	18.44	
	7	21	342	5.70	16.29	
	8	3	58	0.97	19.33	
	9	10	211	3.52	21.10	
	10	17	289	4.82	17.00	
	11	25	396	6.60	15.84	
	12	7	146	2.43	20.86	
	13	13	293	4.88	22.54	
	14	20	405	6.75	20.25	Item #2 had a problem where all PubMed Related Citations recommended contained zero MH tags from which to pull info from. This caused an error to propagate back to ind_batch where it exited. Problem fixed.
	15	8	96	1.60	12.00	
	16	20	345	5.75	17.25	
	17	6	122	2.03	20.33	
	18	12	283	4.72	23.58	
	19	21	732	12.20	34.86	
	20	13	310	5.17	23.85	
	21	13	370	6.17	28.46	
	22	25	432	7.20	17.28	
	23	16	208	3.47	13.00	
	24	5	177	2.95	35.40	
	25	7	112	1.87	16.00	
	26	7	165	2.75	23.57	
		363	7,147		19.69	
Avg:		13.96	275	4.58		

Detailed Results of Daily Batch Jobs between MARS and MTI

Date	Batch #	Item Count	Process Time (in secs.)	Process Time (in mins.)	Avg. Time Per Item	Notes
7/11/2001	1	16	429	7.15	26.81	
	2	12	300	5.00	25.00	
	3	10	261	4.35	26.10	
	4	10	228	3.80	22.80	
	5	13	282	4.70	21.69	
	6	7	140	2.33	20.00	
	7	16	310	5.17	19.38	
		84	1,950		23.21	
	Avg:	12.00	279	4.64		

Detailed Results of Daily Batch Jobs between MARS and MTI

Date	Batch #	Item Count	Process Time (in secs.)	Process Time (in mins.)	Avg. Time Per Item	Notes
7/12/2001	1	6	122	2.03	20.33	
	2	48	1022	17.03	21.29	
	3	38	662	11.03	17.42	
	4	8	202	3.37	25.25	
	5	34	510	8.50	15.00	
	6	5	77	1.28	15.40	
	7	48	395	6.58	8.23	
	8	20	664	11.07	33.20	
	9	17	259	4.32	15.24	
	10	13	211	3.52	16.23	
	11	15	226	3.77	15.07	
	12	10	192	3.20	19.20	
	13	15	220	3.67	14.67	
	14	54	813	13.55	15.06	
	15	28	495	8.25	17.68	
	16	16	319	5.32	19.94	
	17	22	537	8.95	24.41	
	18	25	273	4.55	10.92	
	19	17	290	4.83	17.06	
	20	7	121	2.02	17.29	
	21	9	165	2.75	18.33	
	22	13	230	3.83	17.69	
	23	5	100	1.67	20.00	
	24	6	128	2.13	21.33	
	25	2	40	0.67	20.00	
	26	8	128	2.13	16.00	
	27	18	321	5.35	17.83	
	28	10	283	4.72	28.30	
	29	2	43	0.72	21.50	
	30	9	133	2.22	14.78	
	31	5	93	1.55	18.60	
	32	14	331	5.52	23.64	
	33	3	68	1.13	22.67	
	34	21	502	8.37	23.90	
	35	2	60	1.00	30.00	
	36	29	522	8.70	18.00	
	37	10	259	4.32	25.90	
	38	36	523	8.72	14.53	
	39	15	282	4.70	18.80	
	40	13	377	6.28	29.00	
	41	18	409	6.82	22.72	
	42	5	84	1.40	16.80	
	43	22	452	7.53	20.55	
		721	13,143		18.23	
	Avg:	16.77	306	5.09		