Assessing the quality of automatic classification of NLM customers’ requests and corresponding automatically generated responses to customers’ requests

**Kate Masterton, Associate Fellow 2013-14**

**Project Sponsors**

Terry Ahmed (RWS)

Dina Demner-Fushman (LHC)

**Additional Project Team Members**

Kirk Roberts (LHC)

Halil Kilicoglu (LHC)

Marcelo Fiszman (LHC)

Ron Gordner (RWS)

Lori Klein (RWS)

Selvin Selvaraj (OCCS)

Karen Kraly (OCCS)

Contents

[Introduction 3](#_Toc397083082)

[Terms Used 4](#_Toc397083083)

[Analysis of PubMed citation correction request classification and automatically generated responses 5](#_Toc397083084)

[Access Datasheet 5](#_Toc397083085)

[Reports 5](#_Toc397083086)

[Quality Control workflow 6](#_Toc397083087)

[NCBI Form 6](#_Toc397083088)

[Consumer health questions and automatically generated responses 7](#_Toc397083089)

[Annotation Tasks 7](#_Toc397083090)

[Reports 8](#_Toc397083091)

[Improving CRC Responses 9](#_Toc397083092)

[Analysis of other Siebel product requests 10](#_Toc397083093)

[Survey of Clinicaltrials.gov requests from Siebel 11](#_Toc397083094)

[Survey of Drug/Product requests from Siebel 11](#_Toc397083095)

# Introduction

The National Library of Medicine (NLM) receives up to 100,000 customers’ requests per year. These requests are diverse and cover topics including indexing policies, registering for clinical trials, and licensing of NLM data. The requests can be submitted by users of NLM products directly from NLM webpages, such as MedlinePlus, PubMed, or DailyMed via a “contact us” form. In addition, users can email NLM Customer Service directly.

NLM Customer Service responds to requests with a stock reply, a tailored stock reply, or a researched answer. Responding to the request typically takes 4-10 minutes and as a result, it costs 8-11 dollars per question to respond. Because of the large volume and the associated cost of responding to requests, NLM has developed and implemented a prototype system to aid in automatically answering requests. It is hoped that such a system can eventually reduce the workload of the Customer Service team and allow NLM to respond to customers more quickly.

The prototype system is referred to as the Customer Request Classifier (CRC). Because a significant portion of Customer Service requests are for changes to MEDLINE/PubMed citations, and because these requests are handled with stock replies, the CRC development team used these requests as a starting point. CRC classifies incoming requests by the type of the request. If the CRC labels a request as a PubMed Citation request, it retrieves the citations listed in the request, checks their status and prepares an appropriate stock reply. Before deploying the system into production, there is a need to test the quality of the automatic classification of requests and corresponding automatically generated answers. The primary task is to assess the quality of the classification and answers.

In addition, the CRC development team has an interest in attempting to classify and generate responses for Reference Requests. This is a more complicated and challenging task, but nonetheless Reference Questions and automatically generated responses were evaluated along with the PubMed citation correction requests.

Finally, there may be other types of requests received routinely by NLM that could be automatically handled by CRC.

The following report outlines the activities of the Associate Fellow (Kate Masterton) throughout a year working on the CRC project. The files associated with this project have been saved in a zip file and posted along with this report (MastertonCRC\_files.zip). The file path within the zip file is listed before every file name for ease of navigation.

# Terms Used

CRC – Customer Request Classifier

Siebel – the system used by NLM Customer Service to manage, organize, and respond to all requests sent to NLM (via web form, email, phone, etc.).

SiebelQA – the Siebel test system used by the CRC development team. SiebelQA only receives requests from NLM web forms.

Siebel Production – the Siebel production system used by NLM Customer Service

Quality Control of NLM Databases – the category label for PubMed citation correction requests used in Siebel

Consumer Health Questions – these are the types of questions we would like CRC to handle one day. They are requests for information about a known disease, condition, treatment, etc. from a member of the public.

Example: I have suffered Ankylosing Spondylitis problem since last 2 years in lower back. so plz guid me properly how to cure this problem?

Example: I get numbness to the body alot what should I do

Reference Questions – a label used for customer requests in Siebel. This label applies to a very broad range of reference questions, including ones that we would consider consumer health questions, in addition to many other subcategories.

# Analysis of PubMed citation correction request classification and automatically generated responses

## Access Datasheet

A datasheet in Microsoft Access was used to track requests from SiebelQA. The following request types were tracked in the datasheet:

CRC - indicates that CRC used the correct reply when responding to a request

CRC Error - indicates that CRC did not use the correct reply

CRC Misfire - indicates that CRC tried to answer a request it shouldn’t have

CRC Modified – indicates that CRC would have been correct with slight modifications

CRC Missed - indicates that CRC should have tried to respond to a request but did not

**Outcome**: The Access datasheet was used to generate reports to summarize CRC performance.

## Reports

Using the Access datasheet, monthly reports about SiebelQA performance were compiled. These reports were presented to the CRC Development team, Customer Service, and NLM leadership (Dr. Lindberg and Joyce Backus).

 **November**

**PubMed Citation correction requests 🡪 SiebelQA\_November\_report.docx**

**PubMed Citation correction requests 🡪 SiebelQA\_November\_attachements.docx**

 **December**

**PubMed Citation correction requests 🡪 SiebelQA\_December\_report.docx**

 **January**

**PubMed Citation correction requests 🡪 SiebelQA\_January\_report.docx**

 **February**

**PubMed Citation correction requests 🡪 SiebelQA\_February\_report.docx**

**PubMed Citation correction requests 🡪 SiebelQA\_Feb\_CRC\_Errors\_and\_Misfires.docx**

**Outcome**: After reviewing performance data, it was decided to implement this module of CRC in Siebel Production. The Customer Service team is now monitoring system performance. The latest reports on CRC in Siebel Production from Customer Service are:

 **PubMed Citation correction requests 🡪 CRC Classified Findings of 238 incoming.docx**

 **PubMed Citation correction requests 🡪 CRC priority 20140606 meeting.docx**

## Quality Control workflow

In Summer 2014, the CRC development team had access to three summer interns. Two focused on improving classification of requests from the Siebel category Quality Control of NLM DB. In order to assist the interns’ tasks, a comprehensive view of the workflow for these requests was required. By communicating with Customer Service, we created the following workflow documents:

**PubMed Citation correction requests 🡪 Quality\_Control\_of\_NLM\_DB\_definitions .docx**

 **PubMed Citation correction requests 🡪 Quality\_Control\_of\_NLM\_DB workflow.png**

**Outcome**: The CRC development team now has a workflow diagram for Quality Control of NLM DB requests. This will help build classification rules for CRC.

NCBI Form

It was noted early in the analysis that CRC performed much better with PubMed citation correction requests when a PMID was supplied by the customer. Currently, the form though which the majority of PubMed citation correction requests are submitted does not have a field for PMID. We explored the possibility of creating a new form that would require a PMID for PubMed citation correction requests. This task requires collaboration between NCBI, Customer Service, BSD (because they handle the PubMed citation correction requests), OCCS, and the CRC development team. The following persons are involved in this task:

Kathi Canese (NCBI)

Dina Demner-Fushman (LHC)

Kate Masterton (Associate Fellow)

Terry Ahmed (RWS)

Ron Gordner (RWS)

Ellen Layman (RWS)

Lou Knecht (BSD)

Sara Tybaert (BSD)

Fran Spina (BSD)

Selvin Selvaraj (OCCS)

By communicating between all stakeholders, several documents have been generated:

**PubMed Citation correction requests 🡪 PubMed form 🡪 InitialFormView.docx** -

Initial mockup of what the form would look like

**PubMed Citation correction requests 🡪 PubMed form 🡪Write to the PubMed Help Desk ideas.docx** - This is the current version of the logic for the form

**PubMed Citation correction requests 🡪 PubMed form 🡪PubMed Customer Service Form Revisions.docx** - Revisions for stock replies provided by the form

**PubMed Citation correction requests 🡪 PubMed form 🡪PubMed Form.docx** - Table view of the types of PubMed citation correction requests and how they are handled

**PubMed Citation correction requests 🡪 PubMed form 🡪AllChanges.docx** – shows some of the other requests the form could handle

**Outcome**: Eventually the final mock up version of the form will be passed to NCBI for evaluation. The final outcome for this task will be a new form for PubMed citation correction requests that requires a PMID.

# Consumer health questions and automatically generated responses

Annotation Tasks

Annotating or “marking up” free text provides training data for CRC. During the course of the year, there were three major annotation tasks for the CRC project.

 **Question Decomposition**

These annotations attempt to break apart free text questions. For example:

Original request:

I have an infant daughter with Coffin Siris Syndrome. I am trying to find information as well as connect with other families who have an affected child.

Decomposed request:

S1: [I have an infant daughter with [Coffin Siris Syndrome]FOCUS .]BACKGROUND(DIAGNOSIS)

S2: [I am trying to [find information as well as connect with other families who have an affected child]COORDINATION .]QUESTION

The questions used for question decomposition came from the Genetic and Rare Diseases Information Center or GARD (not from Siebel). We annotated 1,467 multi-sentence questions. For more information about this task, see the following documents prepared by Kirk Roberts:

**Consumer health questions 🡪 annotation docs 🡪 qdecomp\_guideline.pdf** – Guidelines for question decomposition annotation

**Consumer health questions 🡪 annotation docs 🡪 qdecomp\_paper.pdf**  –

Paper outlining question decomposition annotation

**Consumer health questions 🡪 annotation docs 🡪 LREC 2014 Poster.pptx** –

Poster outlining question decomposition annotation

 **Question Type**

These annotations attempt to classify consumer health questions by type of question. Attempting to provide this classification ultimately should improve question responses. For this task, we used the 1,467 decomposed GARD requests, for a total of 2,937 individual questions. For more information about this task, see the following documents prepared by Kirk Roberts:

**Consumer health questions 🡪 annotation docs 🡪 qtype\_guideline.pdf**  -

Guidelines for question type annotation

**Consumer health questions 🡪 annotation docs 🡪 qtype\_paper.pdf**  -

Paper outlining question type annotation

 **Gold frames for Siebel requests**

These annotations attempt to take actual requests from Siebel and create “gold” frames (the frame is what the eventual response is based on; it is essentially what the question is).

 Sample Gold frame:

Original request: my 31 yr old daughter who has c7 she had meningitidis twice when she was 14 yrs @17 she made a full recovery she is now 4 mts pregnant any advice for us please

Question type: Management

Gold frame: MANAGEMENT for [meningitidis] Associated\_with [pregnant]

Theme string: “meningitidis” Question cue string: “advice”

Predicate string: “advice” Associated with string: “pregnant”

The requests from Siebel are more challenging than the GARD requests. In addition, there are many questions labeled as “Reference Questions” in Siebel that are not what we consider consumer health questions, which is really the focus of this task. For example, one of the subcategories of “Reference Questions” is “Patient Records,” which are questions about an electronic health record from customers who come to MedlinePlus from MedlinePlus Connect. These requests are handled with stock replies. While it is possible we may want to attempt to classify and handle these in the future, we won’t need frames for them.

**Consumer health questions 🡪 annotation docs 🡪 Annotation decisions.xlsx** –

Outlines how many of the requests labeled as “Reference Questions in Siebel Production we would want to annotate for our purposes. The ratio is low (37 out of 201).

**Outcome**: The annotations provide training data for CRC. Initial experiments show that so far the annotations have improved CRC performance. More testing and annotating is necessary in the immediate future.

## Reports

Reports of CRC performance with consumer health questions illustrate how we have seen CRC behaving in SiebelQA by highlighting sample requests and responses. Here are sample reports generated by Kate:

**Consumer health questions 🡪 March Response tables.docx**

This document shows side by side CRC responses from SiebelQA and Customer Service responses from Siebel Production

**Consumer health questions 🡪 March Ref Missed and Misfire.xlsx**

This document shows the types of requests in SiebelQA that CRC did not try to answer when it should have (CRC Missed) or tried to answer when it should not have (CRC Misfire)

**Consumer health questions 🡪 March Good Responses.docx**

This shows some of the more promising responses

**Consumer health questions 🡪 FollowUpQuestions\_04\_2014 .docx**

This document highlights some of the types of consumer health questions that we would need additional information to answer (so we would need to “follow up” to answer them)

**Outcome**: These reports help us identify through examples what CRC is doing well and what needs more work. They also highlight questions we need to answer about how to proceed with development.

## Improving CRC Responses

Currently, CRC only pulls content for responses from the MedlinePlus A.D.A.M Encyclopedia, Genetic Home Reference (GHR), and NCBI Gene Reviews. It is hypothesized that increasing the number of resources available for CRC could improve automatic responses.

**Consumer health questions 🡪 2\_12\_14\_Questions w-comments.docx**

Illustrates how some customer requests could be better answered with material outside of the current CRC response corpus

**Consumer health questions 🡪 Source recommendations.docx**

A document prepared for reference for a Summer 2014 intern tasked with building a crawler to enlarge the CRC response corpus

**Outcome**: One of the interns from Summer 2014 built a crawler for several of the sources recommended. The next steps are to index these resources and evaluate if the inclusion of additional resources improve CRC responses.

# Analysis of other Siebel product requests

The Customer Service team uses many categories (75) and subcategories (556) to manually classify incoming Siebel requests. In order to understand if there are other categories that overlap with our current work, I surveyed two of the categories considered potential areas of exploration for us in the future: Clinicaltrials.gov and Drug/Product requests.

| **Product/Category** | **Number (all origins, 01/01/2014-03/31/2014)** | **Product/Category** | **Number (all origins, 01/01/2014-03/31/2014)** |
| --- | --- | --- | --- |
| Document Delivery/ILL | 12958 | WEB Questions-NLM Sites | 37 |
| Reference Questions | 2267 | GHR Genetic Home Reference | 35 |
| Quality Control of NLM DB | 2039 | Non-NLM Products | 30 |
| PubMed | 1838 | Purchasing/Acquisition | 24 |
| MEDLINEplus Spanish | 1431 | SIS | 23 |
| Clinicaltrials.gov | 1278 | LOCATORplus | 20 |
| Drug/Product Questions | 767 | Leasing NLM Databases | 16 |
| MEDLINEplus | 730 | Catalog/Class NLM | 14 |
| Junk Message | 687 | LHC/HPCC | 11 |
| UMLS | 550 | NLM Publications | 11 |
| Duplicate Message | 547 | Training Programs | 11 |
| LinkOut | 545 | Serial Records | 10 |
| Indexing | 304 | Extramural Programs | 9 |
| NCBI | 299 | MEDLINE Data Content | 8 |
| Verifications | 272 | Access NLM Products | 7 |
| NIH Information | 227 | Citing Medicine | 7 |
| NLM General Info | 221 | CustServ Feedback | 6 |
| DOCLINE | 206 | NLM Catalog | 4 |
| Returned Mail | 202 | NNLM | 4 |
| History Questions | 169 | NICHSR Services | 2 |
| PubMed Central | 137 | PubMed Tutorial | 2 |
| LSTRC | 101 | Clinical Alerts | 1 |
| Siebel Support | 96 | Coll Dev Policies | 1 |
| DailyMed | 84 | Comments/Complaints/Sugg. Gen | 1 |
| MeSH | 82 | Customer Service | 1 |
| UNKNOWN | 62 | Digital Repository | 1 |
| NIH Senior Health | 60 | DOCLINE Enhancements | 1 |
| RxNorm | 57 | Newborn Screening Codes | 1 |
| Copyright re NLM Dbases | 55 | NLM DB on Other Systems | 1 |
| Loansome Doc | 44 | Total | 28614 |

## Survey of Clinicaltrials.gov requests from Siebel

**Other requests 🡪 Clinicaltrials.gov\_questionsurvey.docx**

This file outlines the types of requests Customer Service labels as Clinicaltrials.gov and how these requests are responded to.

## Survey of Drug/Product requests from Siebel

**Other requests 🡪 Drug-Product\_questionsurvey .docx**

This file outlines the types of requests Customer Service labels as Drug/Product Questions and how these requests are responded to.

**Outcome**: Recommended that if CRC expands to other requests, it can start with the Drug/Product requests. We are also now in the midst of discussions with the Clinicaltrials.gov team to explore ways automation could potentially help their customer service efforts.