DALHOUSIE UNIVERSITY

Creation and Design of Patient Typing and Comment (PTC) Database

Department of Pathology and Laboratory Medicine: Blood Transfusion Service / Pathology Informatics

John Ku

12/12/2009

In partial fulfillment of the requirements of the Master of Health Informatics Program,

Report of Internship for the period June 16 – December 07, 2009

Acknowledgements

I would like to thank all the co-workers at Pathology Informatics Unit for their support during the period in which this research project was conducted. My profound thanks go to my supervisor, Bryan Crocker, for his encouragement, assistance, and patience in the development of the PTC database. Major thanks go to Craig Winser, Senor Systems Analyst in the IT and Communication department, for providing sound advice and timely assistances in the daily battle with Microsoft Access and Microsoft SQL Server.

I would also like to thank a number of anonymous internet forum users for helping out with some of the technical problems that I faced; and my good friend, Calvin Jien, for answering countless of my newbie programming questions instantaneously over his iPhone.

Finally, I would like to thank Dr. Grace Paterson for her efforts in arranging a number of internship interviews for me, and Dr. Calvino Cheng for providing this valuable learning opportunity to me.

My apologies to those whose help or kindness I have forgotten to or did not mention. Without these individuals, I would not be able to complete my internship completely.

John Ku

Executive Summary

The Patient Typing and Comment (PTC) database was initially proposed by Bryan Crocker and Dr. Calvino Cheng. The division of Pathology Informatics at Capital Health does not have a central database to access patients' information during downtime of the Cerner Pathnet Millennium Laboratory Information System and a database will be needed sooner or later. Currently, all the patients' historical information is combined and recorded in one single text file with limited readability and functions available. The PTC database is designed to address these issues.

The newly developed database provides system users a better method to update patient records, better readability, as well as improved patient search ability. User can search patient record using wildcard search, and retrieve information demographics, medical history, transfusion requirements and all the clinical comments from past encounters. The database provides a base for automatic updates and backup in the future. It will also provide valuable data for data mining and statistical analysis.

There were three major stages to the project:

1. Convert the PTC text file into a format that can be imported into Microsoft Access and.
2. Create a Microsoft Access database for the PTC file.
3. Upsize the Microsoft Access to Microsoft SQL Server, while maintaining a user-friendly frontend interface.

The internship work was originally planned to be performed from June 16 to Sept 16, 2009. However, due to personnel changes and technical problems, the internship was extended to Dec 7, 2009. The internship was a valuable learning experience for the author. Developing a database from scratch is a time-consuming but useful process that will improve PTC usability and potentially reduce transfusion adverse events.

Contents

Introduction

Blood transfusion is a crucial component in many types of medical procedure in which patients lose large amount of blood. More than 24,440 blood transfusions are given each year to patients in Capital Health("Blood Transfusion," 2004), and it is important to manage the patient blood transfusion carefully and systematically to avoid adverse events.

All blood donations in Canada come from healthy donor. Before a patient can receive blood products (whole blood, red cells, plasma, platelets, cryoprecipitate and plasma derivates), a consent form needs to be sign as part of the Consent to Treatment("Blood Transfusion," 2004).
The Transfusion
Laboratory carefully selects and prepares the blood product that is needed by the patient. The laboratory then tests patient's blood with the donated blood, which may take one to two hours to complete, and usually done 48-72 hours in advance of the anticipated transfusion("Blood Transfusion," 2004). This final test ensures that the blood selected for the patient is a cross match with patient's blood type("Blood Transfusion," 2004). As a final check, before a patient is given a unit of blood, two clinicians must check patient's identification armband to ensure the correct blood type is given("Blood Transfusion," 2004). Patient is also being monitored closely during the transfusion to ensure there is no serious reaction to the donated blood("Blood Transfusion," 2004).

Blood types

There are four major blood types: O, A, B, or AB. The blood is also either Rh-positive or Rh-negative. Thus if a person has type A blood, it's either A positive or A negative. The typing depends on the antigens present on the membrane of the red blood cells. A person with blood type A has antigen A present on membrane of red blood cells, along with antibodies B present in body (Figure 1). The blood used in a transfusion must match with patient's blood type, or the antibodies in patient's blood will attack the new blood and causes serious adverse effects.

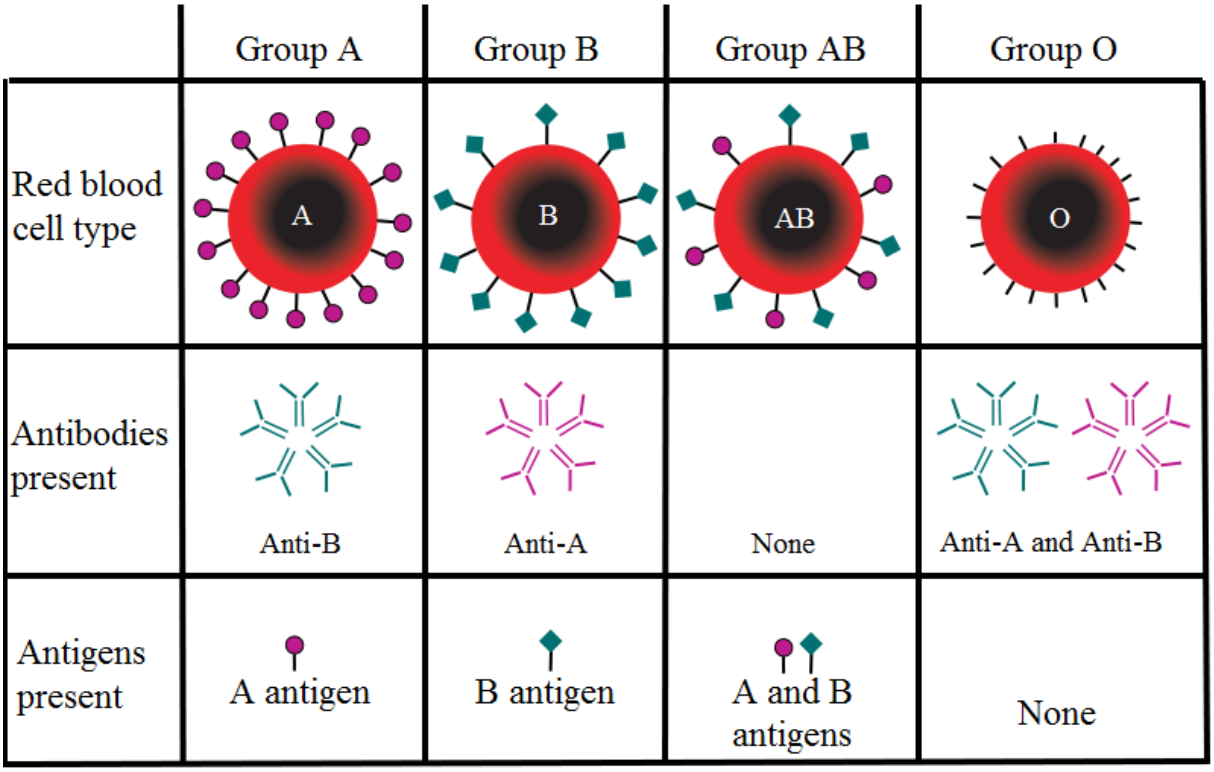

Figure 1: The four major blood types("Blood Type," 2009)

As shown in Figure 1, type O blood contains no antigen, thus it is safe for almost everyone. People who have this blood type are called universal donors. Type O blood is used for emergencies when there's no time to test a person's blood type. On the other hand, patients who have type AB blood are called universal recipients as they do not have any antibody in their blood and can accept any type of blood. In a similar fashion, patients that have Rh-positive blood can get Rh-positive or Rh-negative blood. However, Rh-negative patients can only accept Rh-negative blood. Rh-negative blood is used for emergencies when there's no time to test a person's Rh type.

Aside from the four major blood types, there are subtypes and rare blood types that exist. To date, there are a total of thirty human blood group systems identified by the International Society of Blood Transfusion("Table of blood group antigens within systems," 2008). A numerical terminology system is devised to name the different blood groups by the ISBT (Appendix A).

Description of the Organization

The Department of Pathology and Laboratory Medicine is responsible for managing diagnostic laboratory testing for clinicians and patients of Capital Health. It provides specialized testing and diagnostic consultation for the entire region. There are eight Capital Health laboratories located in Nova Scotia, including two in Halifax: Victoria General Hospital and Halifax Infirmary Hospital("Pathology and Laboratory Medicine," 2009).

Under the Department of Pathology and Laboratory Medicine, the Laboratory Support Services are responsible for Blood Collection Services, Specimen Accessioning and Processing, Central Reporting and Inquiries, and Laboratory Information System (LIS)("Pathology and Laboratory Medicine," 2009). These services are available at every site within Capital health and provide support to all divisions of the laboratories. Over nine million reports are generated annually by the department from five million orderable processes("Pathology and Laboratory Medicine," 2009). All the reports are handled by Cerner PathNet Millennium Laboratory Information System, including the different aspects of specimen collection, handling, ordering, reporting and patient management("Pathology and Laboratory Medicine," 2009).

The Pathology Informatics group is responsible for building, maintaining and troubleshooting the LIS; implementing new functionality and modifying the system to deal with changing laboratory work processes ("Pathology Informatics," 2009).

Description of the Problem

In Blood Transfusion, all the patients' information is captured and handle by Cerner PathNet Millennium Laboratory Information System. An extraction from the information system is performed every week to create a backup file in case the information system goes down. This extract file is known as the Patient Typing and Comments (PTC) Export File. The PTC export file contains all the patient demographic and vital information necessary for the manual process during a system downtime. It is stored on a completely separate server from the Millennium database.

While the PathNet Information System offers clinicians comprehensive, fully integrated technology to automate the operational and managerial sides of the laboratory, with linkage to patient's electronic medical record, the same cannot be said for the PTC Export File. The export file is formatted as a standard text file without any extra function. Any text editors and spreadsheet file can be used in processing or searching the extracted data. However, because it is a standard text, it is very crude and is limited in readability and search functions. There is no security or data integrity in place to protect the data or minimize errors.

Another issue with the PTC Export File is the amount of information it contains. Since it has all patients' information from the first day since the PathNet Information System was put in place, the text file is over several hundred megabytes in size. It can take up to one work day to generate this enormous extract, which puts a strain on the network system. There is also a concern with the ever-growing file size. The size limitation for text file depends on the program that is used to process the file, which is generally very limited. Currently, the PTC Export File has already grown past the limit of Microsoft Notepad and cannot be read by it. The processing of the file relies on UltraEdit (from IDE Computer Solutions). While in theory, UltraEdit has no limit of text file size("UltraEdit," 2009), in practice, it can get very slow when dealing with large file. It also does little to improve readability and search functions.

To address these issues, an idea was brought forth that the PTC Export File needs to be compiled into a database. Cerner recommends importing the file into Microsoft Access database. It will add additional functions as well as reducing the required amount of patients' data extraction. Instead of extracting all information on every patient every week, only an extract of updated data since the previous extract is needed. It can then append the new data into the database.

However, despite of Cerner's recommendations, Cerner does not provide solution on how to implement the system. The Pathology Informatics team does not have any experience in designing a Microsoft Access database. An inquiry was made to the IT department at Capital Health and was told that such project would be placed on queue with a wait time of over 12 months. It became apparent that external help was required in order to expedite the implementation.

Objectives for the Internship

The original objective for the internship was to design an Access database to that can accept the PTC Export File in such a way that it can be run daily or weekly and automatically upload to the access database via server FTP. The new data will then be appended to existing records or create a new patient row if applicable. The database needs to be designed to fit the search requirements of the Blood Transfusion laboratory, with high usability for users with limited database knowledge.

However, under further advice from Dr. Grace Paterson, Assistant Professor in Health Informatics, and Randall Leard, Manager of Application Services for the IT department at Capital Health, it was concluded that Microsoft Access should not be used as the database. Microsoft Access is neither supported nor endorsed at Capital Health. The application has a number of security concerns and should only be use in small office instead or large organization such as Capital Health. Other than security concerns on patient data, the size limit of Microsoft Access is another major concern. The maximum size of the database for Microsoft Access is only 2GB, which will become obsolete in a few years, whereas the maximum size for Microsoft SQL Server is 524, 258TB("SQL Server," 2009). Instead of using Microsoft Access, Microsoft SQL Server was suggested instead. The database design would be designed on Microsoft Access, and then ported over to Microsoft SQL Server via the upsize function while maintaining a Microsoft Access frontend interface.

Challenges

Project Challenges

There were two major challenges that had to be overcome for the completion of the project. The first challenge was the conversion of the PTC Export File. The second challenge was the creation of database. In the early going of the project, after a brief literature search on the two challenges, the second challenge was determined to be the bottleneck of the project. The author had limited experience with database implementation and no experience with neither Microsoft Access nor Microsoft SQL Server. A considerable amount of time was spent on learning the systems, building queries and building small prototype databases using online tutorials.

However, as the author progressed through with the project, it was found that that the biggest challenge was the conversion of the PTC Export File. Upon detail examination of the PTC Export File, it was discovered that due to the structure of the text file, it could not be imported into access without the use of custom programming. The information structure is not consistent throughout a patient's record and it is not delimited in any manner. Figure 2 shows the patient information structure used in the PTC Export File. Each patient record starts with patient's full name and other standard information for the next six lines. If any of the information is missing, it will leave keep the line spacing and allow an empty line. Starting in the eighth nine, however, is the incorporation of tags. Instead of using fixed line structures, start/stop tags are used to indicate the information on transfusion requirements (TR), antibodies (AB), antigens (AG) and comment (COM). For each TR, AB, AG or COM, one line will be allocated for it. If the information is not present, a stop tag will follow directly after a start tag. A detail example of the PTC export is provided in Appendix B.

Patient Full Name
Medical Record Number
Birth Date
Gender
Patient ID Number
ABO/Rh
Phenotype
<START_TR>
<STOP_TR>
<START_AB>
<STOP_AB>
<START_AG>
<STOP_AG>
<START_COM>
<STOP_COM>

Figure 2: Patient information structure in the PTC Export File

It was clear that some form of pre-processing step was needed in order to import the data into any database. With no previous programming experience and a very limited knowledge of PHP and XML, the authors went to numerous online user forums and communities and seek for suggestions. Gathering inputs from over a dozen users, the general consensuses were either some type of parser is needed, or a different extract program was needed. After consulting with

Bryan Crocker, it was not desirable to change the extract program as it would result in major changes to current system, and it should only be seen as a last resort.

After breaking down the issues into smaller problems, the author looked into both XML and PHP language as those were the only programming languages that the author had experience in. At the end the author was able to gather enough help from different random users to put together a PHP code that is able to parse the PTC Export File and convert it into a comma-separated value (csv) file. It started with being able to put together a PHP code that parse the patient information if it was copied and pasted into an input box. It would generate a new text file with uniform structure (Figure 3). It was then improved to generate a csv file that can be imported any spreadsheet or database. Finally, instead of the need of copy and paste patient information into an input box, the PHP code is able to read the information directly from the PTC Export File, which put an end to the conversion challenge. At the time of this report, the conversion can be performed by placing the PTC Extract File just needs into a specific folder and running the PHP file using a web browser. A resulting csv file will be produced in the same folder. The final PHP code is provided in Appendix C.

```
[name] => COE , DOR CAT (INACTIVE)
[medical_num] =>
[dob] => 29/SEP/1923 00:00
[sex] => Female
[id] => 890418
[blood_group] =>
[phenotype] =>
[tr] =>
[ab] =>
[ag] =>
[com] =>
```

Figure 3: Uniform structure of patient information

Once the conversion problem was solved, the creation of the database was done with relative ease. There were some minor issues with the SQL queries, but they were solved with the help of Craig Winser, Senior Systems Analyst of the IT department. Screenshots of the Microsoft Access frontend interface are provided in Figure 4 and Figure 5.

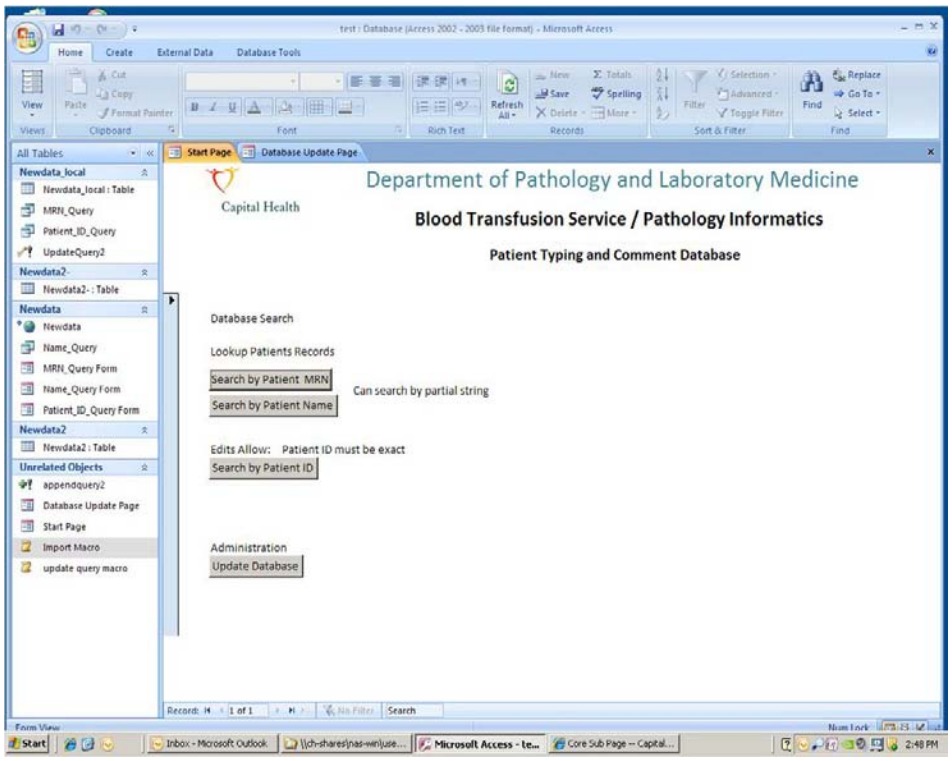

Figure 4: Patient search page

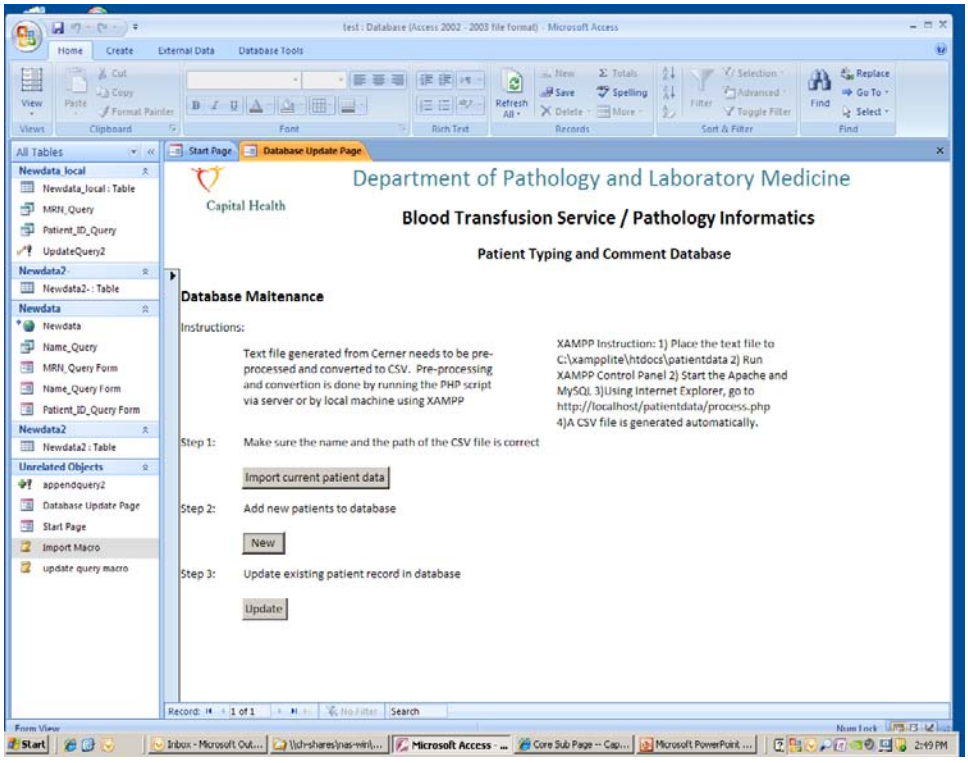

Figure 5: Database maintenance page

Unexpected Issues

There were a number of unexpected issues that occurred during the internship. Eventually these issues were either solved or detours were made to accommodate the issue, at the cost of delaying project completion. There were two areas where problems arose from: personnel and technical issues.

Personnel

There were two unexpected personnel changes that occurred which had negative impact on the project timeline. Firstly, there was supposed to be a resource person from Blood Transfusion that the author was suppose to work closely with. The resource person had in depth knowledge on the PTC export process and would have helped the author in completing the difficult task.

Secondly, before taking on the project, the author recruited another computer science student to help with the technical work required on the project. The computer science student would assist the author in the SQL queries as well as any coding that might be needed in the project. However, three weeks into the project, the student withdrew from the position citing personal issue and her lack of experience in the project area. This forced the author to take on an unexpected load and a steep learning curve for the programming required in the conversion of the PTC Export File. Without knowing the amount of programming needed to parse the PTC Export File, the author decided to proceed on the project alone. Upon review, the project would probably have been finished on time if another programmer was hired.

Technical Issues

Other than personnel problems, the author encountered a number of IT related issues. Since the IT department controls all the IT requests and problems within Capital Health, all the requests are put on a queue and prioritized according to urgency. While it is understandable the importance to maintain centralized control, it slows down the progress of the project significantly. For example, it took two weeks for the author to gain computer access, another week to get Microsoft Access 2003. It took nearly three months before the requested Microsoft SQL Server was installed! Multiple tracking inquires had been made and were unanswered due to personnel on vacation. It wasn't until September 23 that the Microsoft SQL Server was installed. In an internship that was supposed to complete on September 16, it was clear that the business process needed some amendment.

With the problems of lacking resources, obsolete software and aging hardware, the author chose to use his personal computer to press on with the project. Through the Microsoft Developer's Network Academic Alliance (MSDNAA), the author was able to obtain the Microsoft SQL Server 2008. Together with author's own version of Microsoft Access 2007, the author began familiarize himself with the software.

After a draft version of the database was built, a problem arose when the author tried to upsize the Microsoft Access 07 database to Microsoft SQL Server 2008. The upsize process resulted from an error. It was later found that it was a bug from the Microsoft SQL Server 2008 that had not been resolved by Microsoft. The author had to go back and work with the problematic lab computer. With the help of Craig Winser, an evaluation version of Microsoft SQL Server 2005 was installed, which allowed the author to continue the development of the database. While the upsize process from Microsoft Access 2003 to Microsoft SQL Server 2005 worked in the lab computer, it was not able to import any type of file (csv or txt). The program would crash everything an import occurred. This problem did not occur on author's own computer, thus it was suspected that a corruption might have occurred with the program. Microsoft Access 2007 was then installed, hoping that the problem would be resolved – it didn't. The author suspected the lab computer was due for a complete clean up and would require a clean install of every program. Such a move would again be required queuing from IT and take a considerable of time. Since the lab was in a process of replacing the aging computers with latest computers, the project was temporarily put on halt for three weeks.

The breakthrough came with the installation of new computer systems in the lab. All the previous issues of importing and random crashes were resolved. It further took a week to reinstall the different software again, but the project was back on track once again.

Conclusion/Future Works

After extending the internship for three more months, the author was able to create a PTC database with increased usability and search functions of patient information over the PTC Export File. With the new database, it is not necessary to extract all information on every patient each week. Instead, only the new cases need to be extracted and update to the PTC database. By using a Microsoft Access frontend with Microsoft SQL Server backend database, it provides user

friendly information system with advance security and functions. It can be schedule to have automatic updates via the server commands. Currently, the database has not gone live yet, and is situated on a virtual server on the local computer.

The internship at Pathology Informatics was a valuable experience that stretched the author's limit beyond imagination. It allowed the author to apply the knowledge and skills obtained through the Health Informatics program and elaborated on the different teaching points. Despite all the frustrations with vendor software issues, it reflected situations that will happen in the real, where the unexpected do happens. It is never a smooth sailing in the real world, and this internship provided a good taste of it. In this instance, perseverance and teamwork contributed to the success of the project.

Next stage for the project will involve moving the database from the virtual server on the local machine to the real server at Capital Health. Large scale testing by real users can occur and comments can be collected to improve the database.

References

Blood Transfusion. (2004). *Capital Health* Retrieved December 09, 2009, from www.cdha.nshealth.ca/patientinformation/nshealthnet/0647.pdf

Blood Type. (2009). Retrieved Dec 09, 2009, from http://upload.wikimedia.org/wikipedia/commons/3/32/ABO_blood_type.svg

Pathology and Laboratory Medicine. (2009). Retrieved December 09, 2009, from http://www.cdha.nshealth.ca/default.aspx?Page=SubPage&category.Categories.1=62&centerContent.Id.0=49104

Pathology Informatics. (2009). Retrieved December 09, 2009, from http://www.cdha.nshealth.ca/default.aspx?page=SubPage&centerContent.Id.0=2516&category.Categories.1=62

SQL Server. (2009). Retrieved Dec 09, 2009, from http://www.microsoft.com/sqlserver/2008

Table of blood group antigens within systems. (2008, 2008). Retrieved December 09, 2009, from http://ibgrl.blood.co.uk/ISBT%20Pages/ISBT%20Terminology%20Pages/Table%20of%20blood%20group%20antigens%20within%20systems.htm

UltraEdit. (2009). Retrieved Dec 09, 2009, from http://www.ultraedit.com/products/ultraedit/ultraedit_tour.html

Appendix A: Table of blood group antigens within systems("Table of blood group antigens within systems," 2008)

... = obsolete ; provisional numbers are in blue

System		Antigen Number											
		1	2	3	4	5	6	7	8	9	10	11	12
1	ABO	A	B	A,B	A1	...							
2	MNS	M	N	S	s	U	He	Mi^a	M^c	Vw	Mur	M^g	Vr
3	P	P1	...	...									
4	RH	D	C	E	c	e	f	Ce	C^w	C^x	V	E^w	G
5	LU	Lu^a	Lu^b	Lu3	Lu4	Lu5	Lu6	Lu7	Lu8	Lu9	...	Lu11	Lu12
6	KEL	K	k	Kp^a	Kp^b	Ku	Js^a	Js^b	...	...	Ul^a	K11	K12
7	LE	Le^a	Le^b	Le^{ab}	Le^{bH}	ALe^b	BLe^b						
8	FY	Fy^a	Fy^b	Fy3	Fy4	Fy5	Fy6						
9	JK	Jk^a	Jk^b	Jk3									
10	DI	Di^a	Di^b	Wr^a	Wr^b	Wd^a	Rb^a	WARR	ELO	Wu	Bp^a	Mo^a	Hg^a
		1	2	3	4	5	6	7	8	9	10	11	12
11	YT	Yt^a	Yt^b										
12	XG	Xg^a	CD99										
13	SC	Sc1	Sc2	Sc3	Rd	STAR	SCER	SCAN					
14	DO	Do^a	Do^b	Gy^a	Hy	Jo^a	DOYA						
15	CO	Co^a	Co^b	Co3									
16	LW	...	...	...	...	LW^a	LW^{ab}	LW^b					
17	CH/RG	Ch1	Ch2	Ch3	Ch4	Ch5	Ch6	WH				Rg1	Rg2
18	H	H											
19	XK	Kx											
20	GE	...	Ge2	Ge3	Ge4	Wb	Ls^a	An^a	Dh^a	GEIS			
		1	2	3	4	5	6	7	8	9	10	11	12
21	CROM	Cr^a	Tc^a	Tc^b	Tc^c	Dr^a	Es^a	IFC	WES^a	WES^b	UMC	GUTI	SERF
22	KN	Kn^a	Kn^b	McC^a	SI1	Yk^a	McC^b	SI2	SI3	KCAM			
23	IN	In^a	In^b	INFI	INJA								
24	OK	Ok^a											
25	RAPH	MER2											
26	JMH	JMH	JMHK	JMHL	JMHG	JMHM							
27	I	I											

28	GLOB	P											
29	GIL	GIL											
30	RHAG	Duclos	Ol^{a}	Duclos-like									

System		Antigen number											
		13	14	15	16	17	18	19	20	21	22	23	24
2	MNS	M^{e}	Mt^{a}	St^{a}	Ri^{a}	Cl^{a}	Ny^{a}	Hut	Hil	M^{v}	Far	s^{D}	Mit
4	RH	...	...	...	...	Hr$_{o}$	Hr	hr^{S}	VS	C^{G}	CE	D^{w}	...
5	LU	Lu13	Lu14	...	Lu16	Lu17	Au^{a}	Au^{b}	Lu20	Lu21			
6	KEL	K13	K14	...	K16	K17	K18	K19	Km	Kp^{c}	K22	K23	K24
10	DI	Vg^{a}	Sw^{a}	BOW	NFLD	Jn^{a}	KREP	Tr^{a}	Fr^{a}	SW1			
21	CROM	ZENA	CROV	CRAM									

System		Antigen number										
		25	26	27	28	29	30	31	32	33	34	35
2	MNS	Dantu	Hop	Nob	En^{a}	En^{a}KT	`N'	Or	DANE	TSEN	MINY	MUT
4	RH	...	c-like	cE	hr^{H}	Rh29	Go^{a}	hr^{B}	Rh32	Rh33	Hr^{B}	Rh35
6	KEL	VLAN	TOU	RAZ	VONG	KALT	KTIM	KYO	KUCI	KANT	KASH	

System		Antigen number										
		36	37	38	39	40	41	42	43	44	45	46
2	MNS	SAT	ERIK	Os^{a}	ENEP	ENEH	HAG	ENAV	MARS	ENDA	ENEV	MNTD
4	RH	Be^{a}	Evans	...	Rh39	Tar	Rh41	Rh42	Crawford	Nou	Riv	Sec

System		Antigen Number										
		47	48	49	50	51	52	53	54	55	56	57
4	RH	Dav	JAL	STEM	FPTT	MAR	BARC	JAHK	DAK	LOCR	CENR	CEST

Appendix B: Example of PTC Export File for Two Test Patients

EXPORT TYPE:D, BEGIN DATE:01/JAN/2008 00:01, END DATE:31/JAN/2008 23:59

Test , BTS

0800000532

05/MAY/1945 00:00

Female

858327

AB Pos

<START_TR>

CMV Neg Req'd

IgA Deficient

Irradiated

Plasma Deplete PLT

HLA Matched PLT

HPA1a Neg PLT

Blood Warmer

Washed RBCs

<STOP_TR>

<START_AB>

<STOP_AB>

<START_AG>

Wr(a+)

S-

P1-

Vel+

Fy(a-)

Yt(a+)

K:-3

c-

M+

C+

E-

c+

e+

Jk(b+)

K:1

Cw-

K:2

s-

<STOP_AG>

<START_COM>

>> OCT/23/07 08:18:00 TRETHEWEYD >>>>>>>>>>>>>>>>> Please fax to:

>> OCT/22/07 09:37:00 [COMBINE] [Combine from person_id 924598 to person_id 858327]

[Following comments for to person_id 924598]

>> OCT/18/07 15:53:00 TRETHEWEYD 2007/10/18 15:52

>> SEP/25/07 14:50:00 TRETHEWEYD this is Joan's example

>> SEP/21/07 14:34:00 TRETHEWEYD this is another comment.

>> JUL/11/07 11:06:00 TRETHEWEYD this is blood bank comment

[Following comments combined from person_id 924598]

>> OCT/22/07 09:14:00 GAETZM trying to figure out autologous

<STOP_COM>

Test , Janet

8000000056

01/JAN/2000 00:00

Female

949303

O Pos

<START_TR>

<STOP_TR>

<START_AB>

Anti-Wra

<STOP_AB>

<START_AG>

<STOP_AG>

<START_COM>

<STOP_COM>

Appendix C: PHP Code for the PTC Parser

```
<?php
		$source_filename = 'newdata.txt';
		$patient = array(
			'name' => '',
			'medical_num' => '',
			'dob' => '',
			'sex' => '',
			'id' => '',
			'blood_group' => '',
			'phenotype' => '',
			'tr' => '',
			'ab' => '',
			'ag' => '',
			'com' => ''
		);
		//GET ALL DATA
		$handle = fopen($source_filename, 'rb');
		$data = fread($handle, filesize($source_filename));
		fclose($handle);
		//WRITE OUT RESULTS
		$handle = fopen($source_filename . '.csv', 'wb');
		//CLEAN UP DATA
		$data = str_replace('"', '""', $data);
		$data = str_replace("\r", '', $data);
		//PROCESS DATA HERE
		$data = explode("\n", $data);
		array_shift($data); //REMOVE FIRST LINE
```

```
while (sizeof($data)) {
        $line = trim(array_shift($data));
        switch ($line) {
                default:
                        $patient['name']       = $line;
                        $patient['medical_num'] = trim(array_shift($data));
                        $patient['dob']        = trim(array_shift($data));
                        $patient['sex']        = trim(array_shift($data));
                        $patient['id']         = trim(array_shift($data));
                        $patient['blood_group'] = trim(array_shift($data));
                        $patient['phenotype']   = trim(array_shift($data));
                        break;
                case '<START_TR>':
                        $line = trim(array_shift($data));
                        $patient['tr']  = '';
                        while ($data && $line != '<STOP_TR>')  {
                                $patient['tr']  .= $line . "\n";
                                $line = trim(array_shift($data));
                        }
                        break;
                case '<START_AB>':
                        $line = trim(array_shift($data));
                        $patient['ab']  = '';
                        while ($data && $line != '<STOP_AB>')  {
                                $patient['ab']  .= $line . "\n";
                                $line = trim(array_shift($data));
                        }
                        break;
```

```
                                        case '<START_AG>':
                                                $line = trim(array_shift($data));
                                                $patient['ag']  = '';
                                                while ($data && $line != '<STOP_AG>')  {
                                                        $patient['ag']  .= $line . "\n";
                                                        $line = trim(array_shift($data));
                                                }
                                                break;
                                        case '<START_COM>':
                                                $line = trim(array_shift($data));
                                                $patient['com'] = '';
                                                while ($data && $line != '<STOP_COM>') {
                                                        $patient['com'] .= $line . "\n";
                                                        $line = trim(array_shift($data));
                                                }
                                                //WRITE PATIENT DETAILS INTO THE DATABASE HERE USING
THE $patient ARRAY INFORMATION AS DECLARED ABOVE
                                                echo '<pre>';
                                                echo '<b>PATIENT DETAILS</b><br />';
                                                print_r($patient);
                                                echo '</pre><br /><hr /><br /><br />';
                                                fwrite($handle, '"' . implode('","', $patient) . '"' . "\n");
                                                break;
                                }
                        }
                        fclose($handle);
rename("newdata.txt.csv", "newdata2.csv");
```