PERGAMON

Computers in Biology and Medicine 30 (2000) 225–235

Computers in Biology and Medicine

www.elsevier.com/locate/compbiomed

A Java-based application for differential diagnosis of hematopoietic neoplasms using immunophenotyping by flow cytometry

Andy N.D. Nguyen[a,*], John D. Milam[a], Kathy A. Johnson[b], Eugenio I. Banez[c]

[a]*Department of Pathology and Laboratory Medicine, University of Texas Health Science Center at Houston, Texas, USA*

[b]*Department of Health Informatics, University of Texas Health Science Center at Houston, Texas, USA*

[c]*Department of Pathology, Baylor College of Medicine, Houston, Texas, USA*

Received 4 August 1999; accepted 6 January 2000

Abstract

We describe the implementation of a Java-based application for differential diagnosis of hematopoietic neoplasms using immunophenotyping by flow cytometry. The current version of this Java applet includes the knowledge-base for 33 hematopoietic neoplasms and 43 diagnostic immunophenotyping markers. Java, a new object-oriented computing language, helps facilitate development of this applet, a platform-independent module that can be implemented on the World Wide Web. As the Web rapidly becomes more accessible to users around the world, Web-based software may eventually form the core of decision-support systems in clinical settings. Java-based applications, such as the one described in this paper, are expected to contribute significantly in this area.

Keywords: Hematopoietic neoplasms; Immunophenotyping; Flow cytometry; Java; World Wide Web

1. Introduction

Immunophenotyping has become one of the essential methods for proper classification of

* Corresponding author. Tel.: +1-713-500-5337; fax: +1-713-500-0730.
E-mail address: nguyen@casper.med.uth.tmc.edu (A.N.D. Nguyen).

0010-4825/00/$ - see front matter
PII: S0010-4825(00)00007-X

hematopoietic neoplasms. Flow cytometry used in immunophenotyping has no doubt added a new dimension to the diagnosis of leukemia and lymphoma [1,2]. A wide range of monoclonal antibodies is currently available to recognize various hematopoietic cells based on their surface and cytoplasmic antigens [1–4]. Leukemic and lymphoma cells cannot usually be detected with a single immunologic marker. Instead, the use of a monoclonal antibody panel consisting of multiple antibodies is required for supporting the provisional diagnosis based on histological findings [1,2]. Since many hematologic neoplasms demonstrate similar patterns of immunophenotyping, their diagnosis often presents a challenge to pathologists who interpret flow cytometric data. This is particularly applicable to practicing pathologists who are not subspecialized in hematopathology as well as pathology residents in training. As the number of immunologic markers used in flow cytometry increases, a systematic approach in interpretation of marker results is also essential for consistent classification of neoplasms [5].

A relational database, CD-Marker, had previously been developed by the authors to teach pathology residents at our institution about the analysis of immunophenotyping results in lymphoma and leukemia [6]. CD-Marker was designed to run on a PC (IBM-compatible personal computer) with Microsoft Windows (Redmond, Washington). To use CD-Marker, users have to install the run-time version of this database in their computer. The requirement for installation limits the number of users for this single-user software.

The World Wide Web (WWW) offers a simple solution to this problem by providing easy access to teaching materials. Existing internet networks across multiple platforms can be used as the medium for software implementation. Users located in any part of the world with internet connection can use a Web browser to get access to teaching materials that reside in centralized web servers.

Java, a new object-oriented language, has recently received considerable attention due to its association with the creation of Web-based software modules [7]. These modules, known as "applets", can be embedded in traditional Web pages and can execute on any computer platform. Java's wide acceptance has been mainly due to its association with the Web and also due to its technical merits. Java was conceived as an object-oriented language, like C++, by James Gosling and colleagues at Sun Microsystems (Palo Alto, California) in the early 1990's. Long before Sun settled on the current trendy name, Java began life as Object Application Kernel (OAK) in 1992. OAK was designed to be embedded within consumer electronics products such as hand-held devices and interactive televisions. With little commercial success, Sun recast OAK as a programming language for the Web called Java and also introduced the WebRunner browser in 1995. It was not until Netscape (Mountain View, California) incorporated Java support in Navigator 2.0 later that year that Java started to get significant attention. The real charm of Java is its run-time environment that can be executed on any computer system. In traditional languages like C++, compilers are used by programmers to generate sets of native instructions in the form of binary files for a specific operating system, such as Windows, Mac, or UNIX (Fig. 1). By contrast, Java codes are compiled into sets of instructions called Java byte-codes (Fig. 2). These Java byte-codes can be executed on any computer system that has a Java virtual machine (VM), which simulates a computer in software mode. This Java VM can run on existing operating systems or it can run on hardware designed only for Java. This cross-platform concept is appealing to developers who write enterprise applications for organizations that use a variety of platforms. Currently, the most

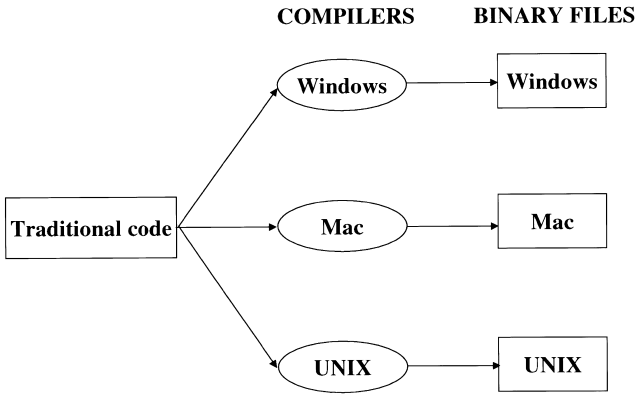

Fig. 1. Traditional compiled programs.

common use of Java is the development of applets deployed on the Web. Java applets ease the implementation of interactive components that users can view in a browser. All the popular browsers (Microsoft Internet Explorer, Netscape Navigator, and Netscape Communicator) are currently Java-enabled.

In this article, we describe the design and implementation of a Java-based program for differential diagnosis of hematopoietic neoplasms using flow cytometry data at our institution, and discuss the potential role of this program in medical education as well as in clinical consultation.

2. Design and results

Web-Marker, a Java-based module for differential diagnosis of hematopoietic neoplasms using flow cytometry data, was implemented in Java using Symantec Visual Cafe 3.0

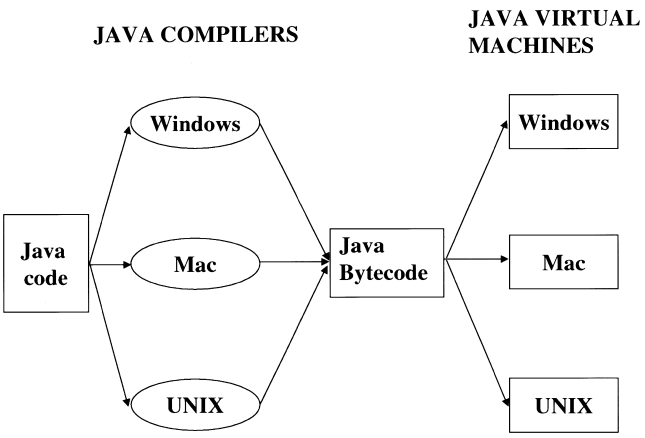

Fig. 2. Java programs.

(Cupertino, California). The codes for the graphic user interface (GUI) and the inference algorithm were compiled into a Java byte-code. This byte-code was embedded into a Web page written in Hypertext Markup Language (HTML) [8]. Web-Marker was installed on a Microsoft Windows NT 4.0 server running Microsoft Internet Information Server 3.0 in the Department of Pathology and Laboratory Medicine, University of Texas-Houston, Medical School.

The core materials in Web-Marker were derived from those developed for the differential diagnostic module of CD-Marker, a Microsoft Access 97 database for personal computers previously designed by the authors [6]. A total of 33 hematopoietic neoplasms (Table 1) and 43 immunologic markers (Table 2) were included in the knowledge-base of CD-Marker. The search engine in this program, written in Microsoft Visual Basic for Application, was based on

Table 1
List of hematopoietic neoplasms in the knowledge-base

Follicular small cleaved cell lymphoma
Mantle cell lymphoma
Large B-cell lymphoma
Mediastinal B-cell lymphoma
B-cell lymphoma, unclassifiable, mixed small and large cells
Lymphoplasmacytoid lymphoma
Marginal cell lymphoma
Burkitt's lymphoma/acute lymphoblastic leukemia, L3
Splenic lymphoma with villous lymphocytes
Peripheral T cell lymphoma
Lymphoblastic lymphoma (T cell)
Thymoma
Chronic lymphocytic leukemia (B cell)/Small lymphocytic lymphoma
Chronic lymphocytic leukemia (T cell)
Prolymphocytic leukemia (B cell)
Hairy cell leukemia
Sezary syndrome/mycosis fungoides
Adult T cell leukemia/lymphoma
Acute lymphoblastic leukemia (T-cell precursor)
Acute myeloblastic leukemia without maturation, M1
Acute myeloblastic leukemia with maturation, M2
Acute promyelocytic leukemia, M3
Acute myelomonocytic leukemia, M4
Acute monoblastic leukemia, M5
Acute erythroleukemia, M6
Acute megakaryoblastic leukemia, M7
Biphenotypic acute leukemia, AML and T-cell ALL
Biphenotypic acute leukemia, AML and B-cell precursor ALL
Large granular lymphocyte leukemia, NK cell
Large granular lymphoproliferative disorder, T cell
Acute lymphoblastic leukemia (Early-B precursor)
Acute lymphoblastic leukemia (CALLA)
Acute lymphoblastic leukemia (Pre-B)

backward-chaining inference [9–11,18]. CD-Marker has been tested with 92 clinical cases from two tertiary medical centers. It ranked the actual diagnosis as one of the top five differential diagnoses in 93% of the cases tested. Further details of the components, inference algorithm, and validation results for CD-Marker have previously been described [6]. Since the same inference algorithm was used in CD-Marker and Web-Marker, the validation results for CD-Marker were also applicable to Web-Marker. Conversion of the materials from the PC-based version to the Java-based version required a complete rewriting of all the PC software codes to accommodate the web medium. The knowledge-base in Web-Marker was coded directly into the Java file. In CD-Marker, by contrast, the knowledge-base was contained in an Access database.

A list of differential diagnoses is provided by Web-Marker with each set of input data. The differential diagnoses have an assigned value of matching factor (MF). The MF value for a neoplasm reflects how well its immunophenotyping pattern matches the marker data in a given case. This factor is defined as:

$$MF = M \times 100/(M + N) \tag{1}$$

where MF is matching factor for a particular neoplasm ($0 \leq MF \leq 100$); M is the number of attributes of a neoplasm that match the input data; and N is the number of attributes of a neoplasm that do not match the input data.

Note that the value of MF, as defined by Eq. (1), only reflects the similarity between a

Table 2
List of immunologic markers in the knowledge-base

CD1	CD41
CD2	CD42
CD3	CD43
CD4	CD45
CD5	CD56
CD7	CD57
CD8	CD61
CD10	CD71
CD11b	CD77
CD11c	CD79a
CD13	CD103
CD14	HLA-DR
CD15	sIg
CD16	cIg
CD19	PC-1
CD20	TdT
CD21	Cytokeratin
CD22	Glycophorin A
CD23	Co-expression of CD5 and CD19
CD24	FMC-7
CD25	
CD33	
CD38	

neoplasm's attributes and the available data. A high value of *MF* for a neoplasm does not exclude the possibility that more data input with increased value of *N* may eventually decrease its *MF* value.

A demonstration of a case with acute promyelocytic leukemia (AML, M3) is illustrative of how Web-Marker can be used in interpretation of immunophenotyping results and how its search mechanism works. Fig. 3 shows the marker data available for the patient sample. Web-Marker attempts to match this set of data with the diagnostic attributes of 33 hematopoietic neoplasms in the knowledge-base. The diagnostic criteria for acute promyelocytic leukemia are as follows [2–4]:

1. The malignant cells are positive for: CD13, CD15, CD33.
2. The malignant cells are negative for: CD2, CD3, CD5, CD7, CD11b, CD14, CD41, CD42, CD61, CD71, HLA-DR, sIg, Glycophorin A.

The available data in the demonstration case matched the following attributes of acute promyelocytic leukemia:

1. Positive for: CD13, CD33.
2. Negative for: CD3, CD5, CD7, CD14, HLA-DR, sIg.

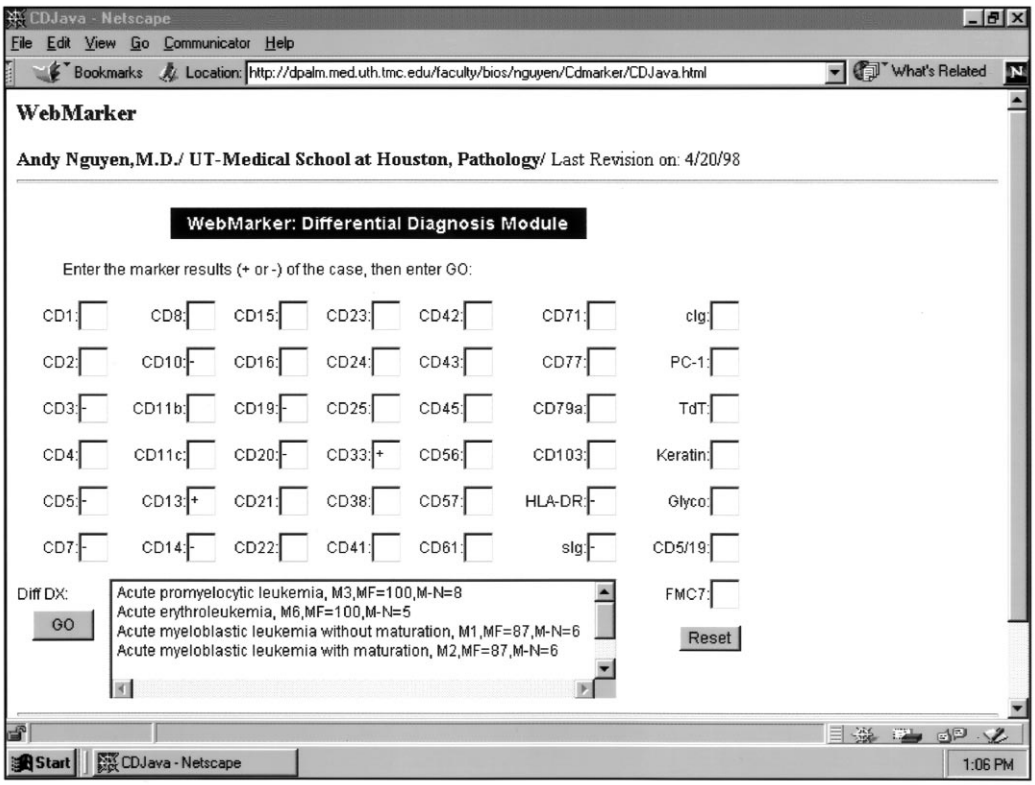

Fig. 3. A demonstration case of acute promyelocytic leukemia.

The total number of attributes of acute promyelocytic leukemia that matched the input data was 8 (two positive results and six negative results). This number was represented by the variable M in Eq. (1) ($M = 8$). None of the attributes of acute promyelocytic leukemia was in conflict with the input data ($N = 0$). Note that the following input data did not have a corresponding attribute for acute promyelocytic leukemia in the knowledge-base: CD10(−), CD19(−), CD20(−). These input data did not have any impact on the ranking of acute promyelocytic leukemia since they were not included as part of the calculation of its MF. Similarly, the following attributes in the knowledge-base without corresponding input data had no impact on the MF value for acute promyelocytic leukemia: CD2(−), CD11b(−), CD15(+), CD41(−), CD42(−), CD61(−), CD71(−), Glycophorin A(−). The intentional exclusion of input data without corresponding attributes in the knowledge-base (or attributes in the knowledge-base without corresponding input data) in calculating MF serves an important purpose of maintaining a flexible design for the knowledge-base as well as for the data input panel. Since different flow cytometry laboratories may utilize different markers in immunophenotyping and various studies on marker patterns of neoplasms have used different marker panels, absolute requirement of certain markers in the interpretation process would be too stringent to yield any reasonable matches [5]. The MF value for acute promyelocytic leukemia at this point was:

$$MF = 8 \times 100/(8 + 0) = 100$$

After Web-Marker calculated the MF value for the remaining 32 hematopoietic neoplasms in the knowledge-base and ranked them accordingly, it listed the following leading diagnoses:

1. Acute promyelocytic leukemia, M3; $MF = 100$.
2. Acute erythroleukemia, M6; $MF = 100$.
3. Acute myeloblastic leukemia without maturation, M1; $MF = 87$.
4. Acute myeloblastic leukemia with maturation, M2; $MF = 87$.

Acute promyelocytic leukemia and acute erythroleukemia had the same MF value ($MF = 100$) with the given input data. A second criteria is used to refine the ranking process for neoplasms with the same MF value. This is the difference between the matched attributes and the unmatched attributes for a neoplasm ($M-N$). With this second criteria, acute promyelocytic leukemia was ranked as the leading diagnosis ($MF = 100$, $M-N = 8$), followed by acute erythroleukemia ($MF = 100$, $M-N = 5$). Fig. 3 shows the list of differential diagnoses with all the calculation results. The search mechanism of going from neoplasms in the database to the input data for the best matches represents a strategy known as backward-chaining search [9–11,18].

This demonstration shows the open-ended format of the data input. The data panel consists of many immunologic markers, some of which may not be part of routine testing in a particular laboratory. Consequently, the actual data input for a case are unlikely to account for all the markers in the data panel. However, the availability of essential data would influence the accuracy of ranking by Web-Marker. Coming back to the demonstration case of acute promyelocytic leukemia, the ranking results would have been different if the result for HLA-DR had not been available. In this scenario (Fig. 4), all the following disorders would have had equal ranking with the same MF value of 100 and ($M-N$) value of 7 [2–4]:

1. Acute myeloblastic leukemia without maturation, M1
2. Acute myeloblastic leukemia with maturation, M2
3. Acute promyelocytic leukemia, M3

The critical role of the interpreting pathologist cannot be overemphasized. Web-Marker is only useful in suggesting a list of differential diagnoses. The pathologist must establish the final diagnosis by correlating the histological findings of the case with the immunophenotyping results.

As shown in the preceding demonstration, Web-Marker was designed with a user-friendly interface. This graphic user interface was arranged such that the sequence of data entry, and display of results should be intuitive to the user. "What If" reruns can be performed whenever the user edits one or more of the input data. This editing may be in the form of changing the value of the data (positive to negative or vice versa), deleting the input data, or adding additional data. The option for reruns enhances the flexibility of Web-Marker in evaluating data that may not be clear-cut, that is, when the marker results are borderline. Without requiring different sets of data entry, the "What If" reruns offer the user a convenient way to consider all the potential diagnoses based on laboratory data that may be subject to errors due to various reasons.

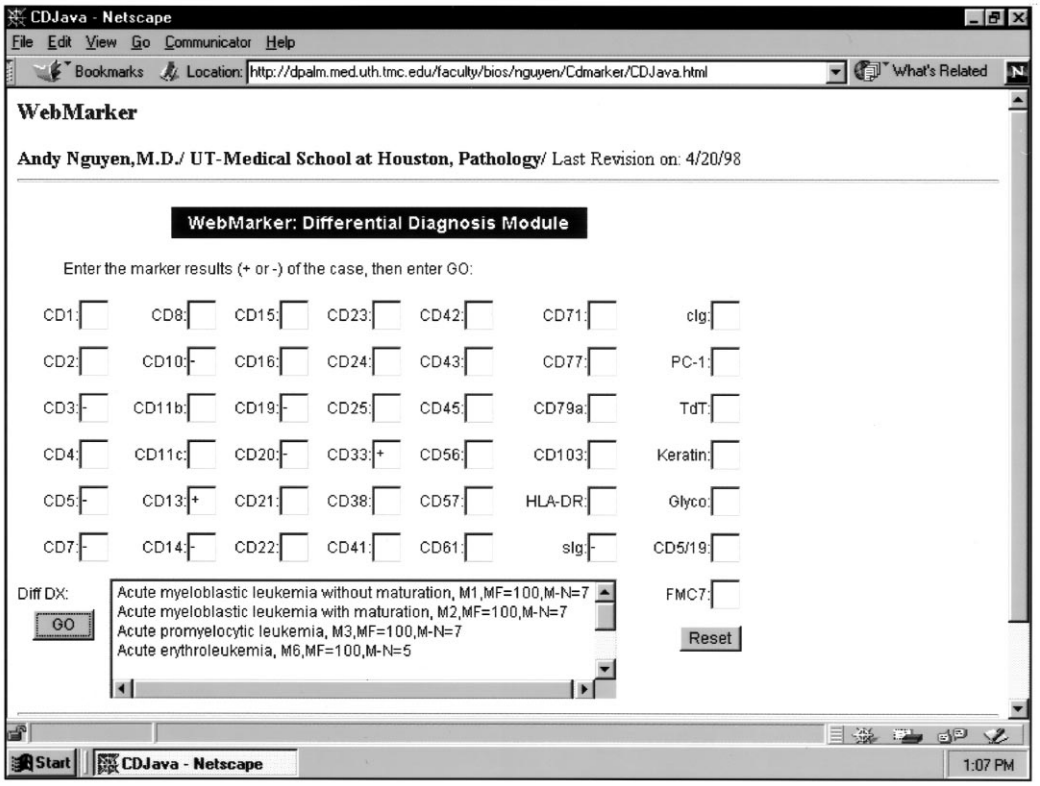

Fig. 4. Output for the demonstration case without data input for HLA-DR.

3. Discussion

We developed a Java-based module for laboratory diagnosis of hematopoietic neoplasms using flow cytometry results. This program has been used as supplemental teaching material for pathology residents on clinical pathology rotations at our institution since 1997. This Web-based program has been received with enthusiasm by our residents.

Unlike traditional models of stand-alone software on a personal computer [6,12,13], the WWW provides users from all over the world and on any computer platform with the same materials located in centralized servers. Updating materials in teaching modules is greatly simplified with this centralization. With the widespread access to WWW resources, many medical educational materials have found their way to the web [14]. Bandwidth limitation has not been a problem for Web-Marker in our experience. The current design of Web-Marker allows for a minimum requirement of hardware and software. A low-end system (486 CPU, 16 megabytes RAM, 14.4 K modem) has been found to be adequate for using Web-Marker.

Despite the utility of Web-Marker, there are certain constraints inherent in its use. Some of these have previously been described for its predecessor, CD-Marker [6]:

1. The user must have a functional knowledge of hematopoietic disorders to be able to use Web-Marker effectively because this program only serves as a search tool to aid the user in making a diagnosis. The technical skills to perform the laboratory procedures and the experience needed to accurately gate the cellular populations are critical in the diagnostic process. Web-Marker can generate a list of differential diagnoses in most cases if adequate data are input. It can not be overemphasized that human judgment is the most important element in finalizing the diagnosis. The number and complexity of hematopoietic neoplasms require that the differential diagnoses suggested by Web-Marker be reviewed before making a final diagnosis. We believe that the pathologist's clinical judgement and the information he or she gathers from Web-Marker should yield an accurate diagnosis.
2. The current version of Web-Marker is deficient in handling some cases of T-cell malignancy due to the difficulty in designing an algorithm for detection of the random loss of T-cell antigens as previously discussed [2–4]. We are currently working on several approaches to alleviate this shortcoming and expect to implement a new technique to handle T-cell malignancies more effectively in future versions of Web-Marker.
3. Web-Marker would not be useful in the diagnosis of neoplasms that traditionally have not been shown to be benefitted from flow cytometric immunophenotyping [2–4]. These include Hodgkin's disease, multiple myeloma, and cases of atypical immunophenotypes.
4. Even though the current versions of all the popular Web browsers support Java, some older versions do not. This will create a problem for users with older versions of Web browsers who attempt to run Web-Marker. Furthermore, the knowledge-base of Web-Marker was coded directly into the Java file. Changing any part of this knowledge-base would require compiling the Java code again. We are currently working on a new version of Web-Marker utilizing Microsoft Active Server platform [15]. This platform allows us to implement dynamic database on the Web that accommodates the use of a "thin" client (a Web browser

without any plug-ins or extensions). The knowledge-base can be easily modified by changing the contents of the database.

4. Summary

In conclusion, we found that Web-Marker provides a convenient, interactive tool to assist clinical personnel in diagnosing hematopoietic neoplasms using flow cytometric data. This Java-based application is available to a large number of users via the WWW. As the Web provides medical students and residents with convenient tools for self-study and clinical decision-support, Web-based education programs may eventually form the core materials for life-long learning by physicians, especially at the point-of-service [8,14,16,17]. Further studies would be needed to determine the most effective ways to utilize this new emerging medium in medical education.

5. Web site URL

Web-Marker can be accessed at the following URL: http://dpalm.med.uth.tmc.edu/faculty/bios/nguyen/nguyen.html

In the event that this URL is changed due to some unexpected reason, please contact the authors by e-mail to obtain the new URL: nguyen@casper.med.uth.tmc.edu.

Acknowledgements

We thank Alex Buraruk, MT (ASCP), for consulting with us on technical matters during the course of this project; Donna Obermeier, BS for many helpful suggestions.

References

[1] D.F. Keren, Flow Cytometry in Clinical Diagnosis, ASCP Press, American Society of Clinical Pathologists, Chicago, 1989.
[2] T. Sun, Color Atlas and Text of Flow Cytometric Analysis of Hematologic Neoplasms, Igaku-Shoin, New York, 1993.
[3] C. Kjeldsberg (Ed.), Practical Diagnosis of Hematologic Disorders, ASCP Press, American Society of Clinical Pathologists, Chicago, 1995.
[4] N.L. Harris, E.S. Jaffe, H. Stein, A revised European–American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group, Blood 84 (5) (1994) 1361–1392.
[5] W. Check, Diagnosing leukemia, lymphoma: when laboratories go with flow analysis, CAP Today November (1998) 1–37.
[6] A.N.D. Nguyen, J.D. Milam, K.A. Johnson, E.I. Banez, A relational database for diagnosis of hematopoietic neoplasms using immunophenotyping by flow cytometry, Am. J. Clin. Pathol. 113 (2000) 95–106.

[7] R.P.C. Rodgers, Java and its future in biomedical computing, J. Am. Med. Inform. Ass. 3 (5) (1996) 303–307.
[8] J.L. Vossler, Developing web-based instruction for the clinical laboratory, Lab. Med. 29 (3) (1998) 167–173.
[9] M. Chignell, K. Parsaye, Expert Systems for Experts, Wiley, New York, 1988.
[10] B.G. Buchanan, E.H. Shortliffe (Eds.), Rule-Based Expert Systems, Addison-Wesley, Reading, Mass, 1984.
[11] E. Turban, in: Expert Systems and Applied Artificial Intelligence, Macmillan, New York, 1992, pp. 73–114.
[12] A.N.D. Nguyen, M. Uthman, XPCOAG: a teaching software for laboratory diagnosis of coagulation disorders, Lab. Med. 25 (6) (1994) 399–401.
[13] A.N.D. Nguyen, E.A. Hartwell, J.D. Milam, A rule-based expert system for laboratory diagnosis of hemoglobin disorders, Arch. Pathol. Lab. Med. 120 (1996) 817–827.
[14] E. Klatt, S.E. Dennis, Web-based pathology education, Arch. Pathol. Lab. Med. 122 (1998) 475–479.
[15] C. Willie, Unlocking Active Server Pages, New Riders, Indianapolis, 1997.
[16] K.D. Horn, D. Sholehvar, J. Nine, J. Gilbertson, C. Hatton, C. Richert, M.J. Becich, Continuing medical education on the World Wide Web: interactive pathology case studies on the internet, Arch. Pathol. Lab. Med. 121 (1997) 641–645.
[17] A.N.D. Nguyen, M.O. Uthman, K.A. Johnson, A Web-Based Teaching Program for Laboratory Diagnosis of Coagulation Disorders. Arch. Pathol. Lab. Med. In press.
[18] E. Turban, in: Expert Systems and Applied Artificial Intelligence, Macmillan, New York, 1992, pp. 435–480.

Andy N.D. Nguyen, Department of Pathology, University of Texas Health Science Center, 6431 Fannin, MSB 2.292 Houston, Texas 77030. MD, 1989, University of Texas Medical Branch, Galveston, Texas; Assistant Professor, Department of Pathology, University of Texas Health Science Center; Hospital affiliation: Lyndon B. Johnson, Memorial Hermann, Houston Texas; Areas of interest: medical informatics, hematopathology, and coagulation.

John D. Milam, Department of Pathology, University of Texas Health Science Center, 6431 Fannin, MSB 2.292 Houston, Texas 77030. MD, 1960, Louisiana State University, School of Medicine, New Orleans; Professor, Department of Pathology, University of Texas Health Science Center; Hospital affiliation: Lyndon B. Johnson, Houston Texas; Areas of interest: transfusion medicine, clinical laboratory management.

Kathy A. Johnson, Department of Health Informatics, School of Allied Health Sciences, University of Texas Health Science Center, 7000 Fannin St., Suite 690, Houston, TX 77030. PhD, 1993, Ohio State University; Assistant Professor, Department of Health Informatics, School of Allied Health Sciences, University of Texas Health Science Center; Courses taught: Knowledge modeling in healthcare, Advanced database concepts in healthcare; Areas of interest: artificial intelligence, intelligent tutoring systems, and cognitive science.

Eugenio I. Banez, Department of Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. MD, 1973, University of the Philippines; Associate Professor, Department of Pathology, Baylor College of Medicine; Hospital affiliation: Ben Taub, Houston Texas; Areas of interest: hematopathology, and flow cytometry.