ELSEVIER

Computer Methods and Programs in Biomedicine 53 (1997) 119–133

Computer Methods and Programs in Biomedicine

TOXOPERT-I: knowledge-based automatic interpretation of serological tests for toxoplasmosis

Sándor Nagy [a,*], Michael Hayde [b], Birgit Panzenböck [b], Klaus-Peter Adlassnig [a], Arnold Pollak [b]

[a] *Department of Medical Computer Sciences, University of Vienna Spitalgasse 23, A-1090, Vienna, Austria*
[b] *Department of Neonatology, University Children's Hospital, Währinger Gürtel 18-20, A-1090, Vienna, Austria*

Received 10 May 1995; received in revised form 10 February 1997; accepted 13 February 1997

Abstract

Primary infection with *Toxoplasma gondii*, a parasite found in most regions of the world, is asymptomatic in more than 80% of cases. However, primary infection with *Toxoplasma gondii* in a pregnant woman might cause fetal infection and severe damage. Most cases do not require treatment. This applies to women without any infection (denoted as seronegative) and women who have acquired the infection before conception (denoted as latent). In contrast, women with postconceptual infection require immediate treatment to prevent or ameliorate fetal infection. We have developed an expert system, called TOXOPERT-I, designed for routine laboratory work, which automatically interprets serological test results of toxoplasma infection. By using the system the clinician can also examine questionable cases by interactively exploring possible results. We used a popular method of designing expert systems applied to medical interpretation and therapy advice, the rule-based one. In order to meet the requirements of automatic interpretation in toxoplasma serology the following characteristics were introduced: the interpretation of sequences of test results, the possibility of excluding inconsistent test results and the adaptability of the knowledge base. A decision graph that covers the different kinds of infections as well as therapy and recommendations for further tests was designed, implemented and was clinically tested by carrying out a retrospective study including 1000 pregnant women. A comparison of TOXOPERT-I and the clinician's interpretations yielded sensitivity and specificity rates of over 99% each. © 1997 Elsevier Science Ireland Ltd.

Keywords: Artificial intelligence; Expert system; Computer-assisted decision making; Decision graph; Serology; Toxoplasmosis

* Corresponding author. Present address, until 19 July 1997, Department of Computer Science and Informatics, Faculty of Science, The University of the Orange Free State, P.O. Box 339 Bloemfontein 9300, Orange Free State, Republic of South Africa. Tel.: +27 51 4012754; fax: +27 51 4474152; e-mail: sandor@wwg3.uovs.ac.za; and from 20 July 1997, Neue-Weltgasse 8/5, A-1130 Vienna, Austria.

PII S0169-2607(97)01813-0

1. Introduction

Before the initiation of obligatory country-wide serological screening for pregnant women in Austria, the incidence rate of prenatal toxoplasma infection was 50–70 per 10 000 births (5–7‰). At present the incidence rate is less than one per 10 000 births (< 0.1‰) [1]. Toxoplasmosis is a disease caused by a parasite called *Toxoplasma gondii* which is widespread all over the world. The distinction between disease and infection is clinically and epidemiologically essential [2]. Human infection is acquired mainly by ingestion of undercooked or raw meat containing viable cysts or vegetables and salads contaminated with mature oocysts (excreted only by members of the cat family). Other means of infection are transplacental transmission or, less commonly, through organ transplantation, blood or leukocyte transfusion and laboratory accidents.

Infection with *Toxoplasma gondii* is usually not symptomatic in the immunocompetent individual, whereas a first infection during pregnancy (denoted as acute infection) causes fetal infection of up to 60% if not treated properly. Clinical manifestations range from early fetal death to asymptomatic, apparently healthy new-borns. The latter however, carry a high risk of severe eye complications (choreoretinitis and blindness) later in life.

If maternal infection is discovered early after its onset, the fetus can be protected from placental parasite transmission by appropriate therapy. This is the reason for a general screening program based on serology which was established in Austria in 1975 [1] and in France in 1978 [3] to detect and treat maternal infections.

The question of whether a toxoplasma infection will cause fetal health problems can be answered by distinguishing between latent infection, which was acquired before conception, and acute infection, which was acquired after conception. Latent maternal infection protects the fetus from infection because sufficient maternal antibodies against *Toxoplasma gondii* are available to prevent placental parasite transmission. If a positive IgG titer prior to conception is present, e.g. a positive screening result from a former pregnancy is available, or if IgG titer remains at low levels, i.e. 1:4–1:256 in follow-up tests, latent infection can be interpreted. However, seronegative women, who are susceptible to infection, must be continually observed during pregnancy.

Our consideration was that interpretation of toxoplasma infection by means of serological test results can be supported by an expert system. Based on several requirements (Section 2.2) that any expert system applied to toxoplasma serology should meet, TOXOPERT-I was developed. This paper describes TOXOPERT-I in Section 2 and Section 3. Section 4 presents samples of typical system runs. Section 5 reports on a retrospective study carried out with 1000 cases and discusses the results achieved. Finally, Section 6 analyzes published alternative approaches.

2. System design considerations

2.1. Medical background

General mass screening for toxoplasmosis has been established in Austria since 1975 as a part of the 'mother–child care program' [1,4]. To meet the demands of that screening program the following two serological tests have been routinely applied in our laboratory.

1. *Sabin-Feldman dye test (DT)*
   The IgG DT is the World Health Organization (WHO) reference IgG test (toxoplasma-specific immunoglobuline G). Thus, positive titers prove and negative titers exclude infection. This test is solely carried out in a few reference laboratories because it is somewhat cumbersome to perform because it requires living parasites. The distinction between latent infection, which protects the fetus, and acute infection is sometimes difficult to make because of the considerable variation of individual immune responses.
2. *Immunosorbent agglutination assay (IgM ISAGA, Bio Mérieux®)*
   The IgM ISAGA is used for the detection of toxoplasma-specific IgM antibodies (immunoglobuline M; macroglobuline) to give additional information on the stage of infection.

Both tests can be described by their idealized titer curve as a result of antibody response to acute toxoplasma infection [5,6] (Fig. 1).

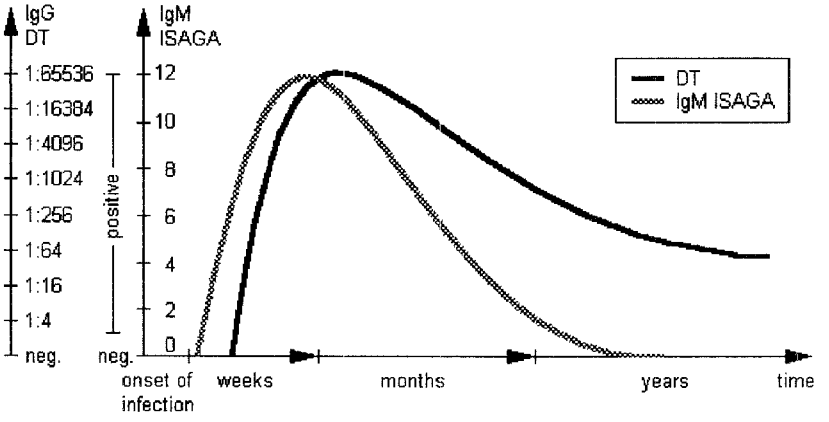

Fig. 1. Idealized titer curve of IgG and IgM antibodies for acute toxoplasma infection. IgM titer rises faster to a maximum level and achieves a negative level after the acute phase. However, it can persist at an individual peak level for several years. IgG titer achieves its maximum level after about nine to twelve weeks and falls after 4–9 months but persists on positive levels for life.

As Fig. 1 shows, DT becomes positive immediately after infection ($\geq$ 1:4), reaches a maximum value (1:65536) after about 9–12 weeks, stays usually at individual peak titer levels for 4–9 months, decreases slowly over the years and persists throughout life at low levels, i.e. between 1:4 and 1:256. By comparison, IgM titer falls rapidly to a negative value, i.e. toxoplasma-specific IgM antibodies are not detectable, but can persist in some individuals for up to 36 months. Therefore, the probability of discovering the early rise in the IgM titer which would indicate an acute infection is small because of the relatively short period of initial titer rise in contrast to the long period of decreasing titer in the latent phase of infection.

To recognize whether or not a patient is infected with *Toxoplasma gondii*, a clinician has to compare the patient's DT titer curve with the general, empirically determined titer curve that is based on follow up serology tests of normal individuals within a population. A negative DT result (no toxoplasma-specific IgG antibodies are detectable, i.e. IgG titer $<$ 1:4) indicates no exposure, no infection and thus, no immune response, in contrast to a positive DT result (IgG titer $\geq$ 1:4). If IgG titer is very high (IgG titer $\geq$ 1:16384), the patient has been infected recently, i.e. up to a few months ago. If DT is low positive (IgG titer between 1:4 and 1:256), then it is not clear if the onset of infection was a long time ago or only recently although the latter is unlikely. Consequently, a few weeks later another test must be performed so as to distinguish between titer rise, which means recent infection, and titer persistence, which means latent infection. Maternal latent infection protects the fetus from toxoplasma infection.

The traditional interpretive process as employed in toxoplasmosis laboratories is depicted in Fig. 2. Two out of the four phases are candidates for complete automation: the second and third phase. TOXOPERT-I has been developed to automatically interpret test results to suggest the respective interpretation. Phase four, the confirmation of the suggested interpretation, should be solely the physician's responsibility. The proposed interpretive process is to modify phase three by performing computer-assisted interpretations by an expert system.

2.2. System requirements

To model the above-mentioned interpretive process of toxoplasma infection an expert system to be used for computer-assisted interpretation of serological test results should meet the following requirements:

1. Sequences of test results have to be analyzed. To compare test results with a titer curve, at least two findings are necessary to detect trends. After each finding the interpretation can be more specific; in other words, the most

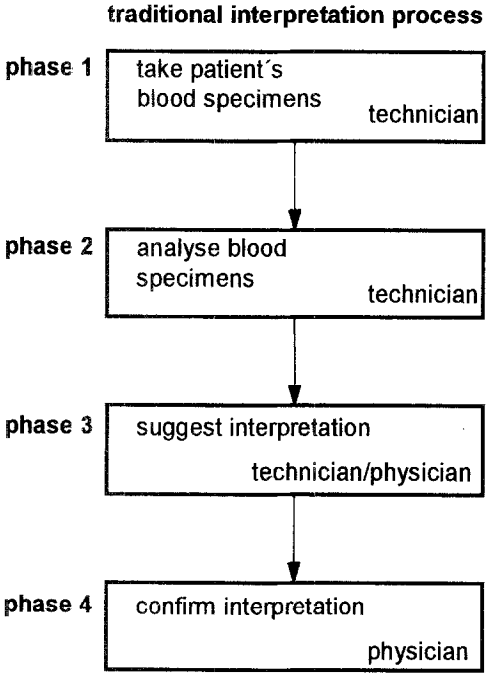

Fig. 2. Traditional interpretive process as carried out in a laboratory without computer assistance.

recent interpretation should explain all findings that have previously been obtained.

2. Inconsistent test results have to be recognized and eliminated accordingly. If a test result turns out to be an outlier—whatever the reason may be—there should be a mechanism that allows the clinician to exclude implausible test results from the set of test results which are used for the serological interpretation.
3. The knowledge base should be open to change. The clinician, who is involved most with the expert system, should be able to understand the structure of the knowledge base and to change parts of it. Furthermore, an explanation module should illustrate how interpretations are derived.
4. Adaptation of the knowledge base to new serological tests should be possible. This requirement incorporates features of an expert system shell. The knowledge base should be adaptable to new serological test methods.

3. Methods

3.1. System description

The general structure of the medical expert system TOXOPERT-I is shown in Fig. 3. It consists of the following components: knowledge base,

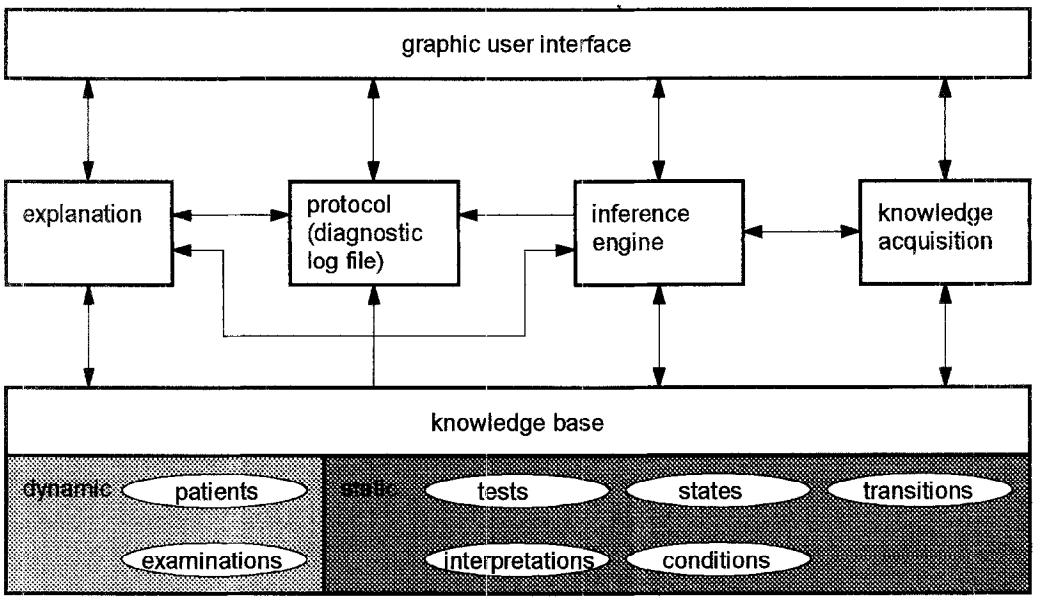

Fig. 3. Block diagram of TOXOPERT-I including the objects of knowledge representation.

inference engine, knowledge acquisition component, explanation module and user interface (Section 4.1). The explanation module was designed as a visualized path on the decision graph and, in addition, a protocol component has the ability to store and make available all interpretations established by TOXOPERT-I. Due to the importance of correct interpretive results, the conditions under which interpretations have been reached must be documented. The knowledge base can be divided into a dynamic part, consisting of the patients' data and their corresponding examinations, and a static part. The static part covers the deductive knowledge, i.e. the decision graph consisting of states and transitions. Every state contains an interpretation including recommendations for further tests. States and transitions can contain a condition. The condition of a state is also called the entry condition because it serves all arriving transitions. Tests are required to formulate examinations and conditions.

3.2. Knowledge representation

Rule-based methodology was chosen to develop TOXOPERT-I. As Fig. 3 shows, knowledge is divided into a dynamic and a static part. In the case of toxoplasma infection an appropriate way to structure the static knowledge (changes or additions are not frequent) is to use a decision graph. As opposed to decision *trees*, decision *graphs* may also include cycles. The decision graph used can be interpreted as a finite state machine (proof can be found in [7]) consisting of states, transitions, and conditions, where each state represents an interpretation. All states have to be connected directly or through other states with the initial state where the interpretive process starts.

In TOXOPERT-I knowledge is represented declaratively and based on propositions in contrast to object-oriented representations. Moreover, the problem-specific semantic is important for the interpretation of the knowledge base by the inference engine. Although representation is not independent from the problem, different medical fields of knowledge with comparable behaviour can be modeled.

The decision graph was designed as a free net structure: each state can be linked with any other state. A result of the free net structure is high inferential efficiency because only states that are connected with the current state are taken into consideration during the inference process. Time is indirectly represented by the chronological evaluation of patient's serological test results.

The incorporated knowledge of the decision graph can be transformed into a rule representation. If that is desired, each possible interpretive path that leads to a certain interpretation must be taken into account. Consequently, the more transitions the decision graph consists of, the more complex the rule format becomes. At present the decision graph consists of 67 states (including the initial state which does not represent an interpretation), 145 transitions and 93 conditions. The present number of different sequences of examinations that can be treated by the decision graph is infinite due to existing cycles. If the cycles are not taken into consideration and an upper boundary of 40 weeks gestational age (GA) is assumed, 247 934 838 different sequences of examinations may be interpreted by the present structure of TOXOPERT-I's knowledge base. The calculation is a result of all different interpretive paths, including all possible test results for each interpretive path where the interpretive path complies with the corresponding conditions.

3.3. Inference engine

The inference engine is realized by deductive reasoning within a decision graph (Section 3) using forward chaining without backtracking. That decision graph can be seen as a finite state machine.

The serological tests are modeled by separate test units (denoted as *tests* in Fig. 3) that are connected with a certain data format. These data formats can be date, time, float, integer and collection. Test results must correspond with the test unit that they belong to. The format *collection* describes a set of ordered items. Each item is an arbitrary text. A test result must be one of those items. Test results of data format *collection* are compared with one another by comparing their

Table 1
Example of patient's test results

Finding/examination	Date	GA in weeks	IgG	Computed IgG trend	IgM
Previous	04/17/1996	10	negative		not tested
1	06/05/1996	18	negative	$\perp$	not tested
2	07/30/1996	26	1:256	$\uparrow\uparrow$	positive
3	08/27/1996	30	1:16384	$\uparrow\uparrow$	positive

GA denotes gestational age in weeks.

indexes (in the following denoted as *ind*(x)) in the set; e.g. if we have a set $IgG = \{negative, 1{:}4, 1{:}16, 1{:}64, 1{:}256, 1{:}1024, 1{:}4096, 1{:}16384, 1{:}65536, > 1{:}65536\}$ then the condition $1{:}1024 \geq 1{:}64$ is true because *ind* $(1{:}1024) = 6$ and *ind* $(1{:}64) = 4$ and $6 \geq 4$ is true.

A simple comparison has the format *val (test)*$\circ$*test result (test, n)*, where *val*(*test*) denotes a value or an item of *test*, e.g. '1:1024' or 'negative', $\circ$ is one of the comparison operators $\{=, \neq, >, \geq, <, \leq\}$, and *test result*(*test*, *n*) denotes the test result of *test* at the *n*-th examination. A sequence of Boolean AND-combined simple comparisons generates a condition. If a patient's examination complies with the condition in the transition (if available) and the corresponding entry condition of the next state, this state becomes the current state and inference continues with following examinations. When the inference procedure stops the interpretation of the last current state is the result of the inference.

3.4. Knowledge acquisition

Knowledge acquisition is indirectly achieved with the help of a knowledge engineer. Data input can be effected directly via user interface or by using ASCII file import and/or export. Currently, the knowledge engineer takes on responsibility for the completeness (all test results generate a serological interpretation) and consistency (no test result can generate contradictory interpretations) of the decision graph.

A short example should illustrate how the decision graph and the inference engine function. Let us assume the following examples of test results (Table 1), where the symbol $\perp$ denotes a constant IgG trend and $\uparrow\uparrow$ denotes a significantly rising IgG trend. In addition, let us assume the following sample knowledge base (conditions are given next to transitions) and the following interpretive path (Fig. 4).

The interpretive process starts with the initial state, denoted as *start*, which becomes the current state. For each examination exactly one transition is evaluated. An examination satisfies a transition if it satisfies the condition of the transition and the entry condition of the state to which the transition is leading. Conditions are separated into these two parts because knowledge representation becomes more efficient and shorter since entry conditions have to be satisfied by all arriving transitions. Thus, redundancies are avoided. Only one transition from the current state can be satisfied by the current examination because of mutual exclusive conditions in one level. The first transition that is satisfied by the current examination is used. The state that the transition is leading to becomes the subsequent current state. The corresponding interpretation becomes the subsequent current interpretation. The process is repeated for the next examination. If no more examinations are available, the interpretive process stops and the interpretation of the most recent current state becomes the result of the deduction.

A deduction of an interpretation can be described or explained by its path through the decision graph. In Fig. 4 the highlighted interpretation *'acute infection confirmed'* is the result of such a deduction. The clinician has now the opportunity to follow the highlighted path through the graph. This kind of explanation facilitates the immediate recognition of well-known

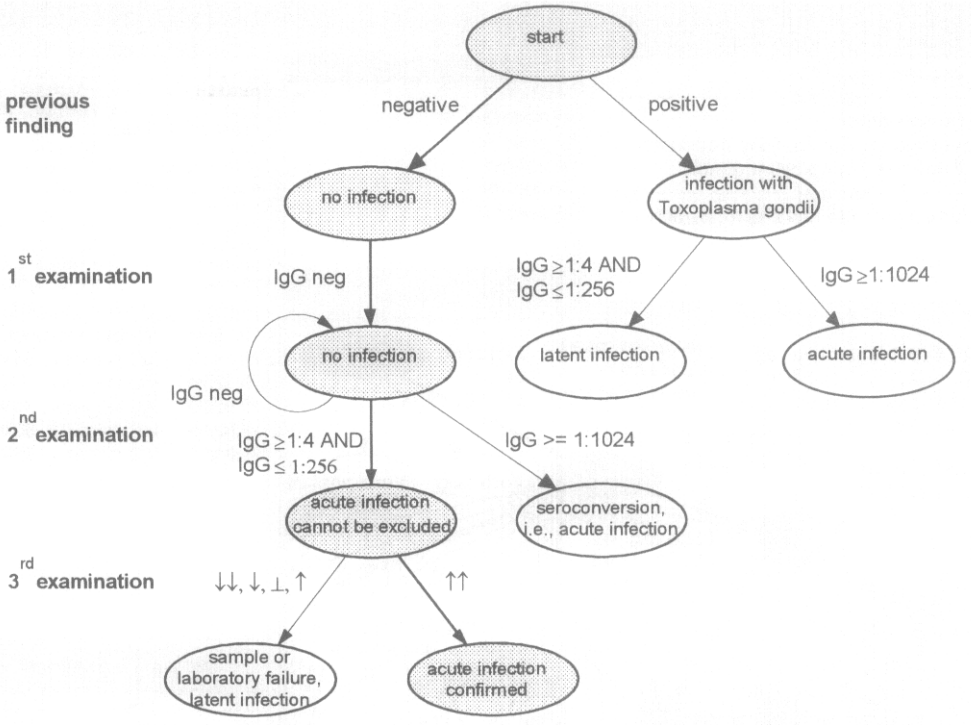

Fig. 4. Interpretive path (visualized interpretation explanation) of an example case. ↑↑ Denotes significant titer rise, ↓↓ denotes significant falling titer and the commas between comparisons refer to OR-concatenation.

interpretations as well as interpretive anomalies. As a result, even clinicians with little experience with computers or expert systems are capable of deciding whether the knowledge base is incorrect or interpretive anomalies occur.

3.5. Explanation module

The necessity of analyzing sequences of examinations and a transparent knowledge base led to the idea of making the knowledge base visual. The explanation module marks all states that are involved in the deduction with a certain color to emphasize the interpretive path. This interpretive path in combination with the protocol component is used to explain how interpretations are derived. The visualization is implemented similarly to Fig. 4 (Figs. 5–7 as well).

3.6. User interface

A graphic user interface (GUI) was designed and implemented to support the user with the information required to use the system and to derive interpretations from serological test results. The aim was to avoid frequent switching among numerous windows to facilitate and expedite routine use. This was achieved by splitting up the main window into six subpanes; each containing a specific part of the knowledge base (Fig. 5). Each subpane can be zoomed to full size and has a corresponding menu that reflects the functions that are applicable to the data within the window. For instance, in the patient data window (on the top right) the menu functions are *add*, *change*, *delete* and *print* the patient data.

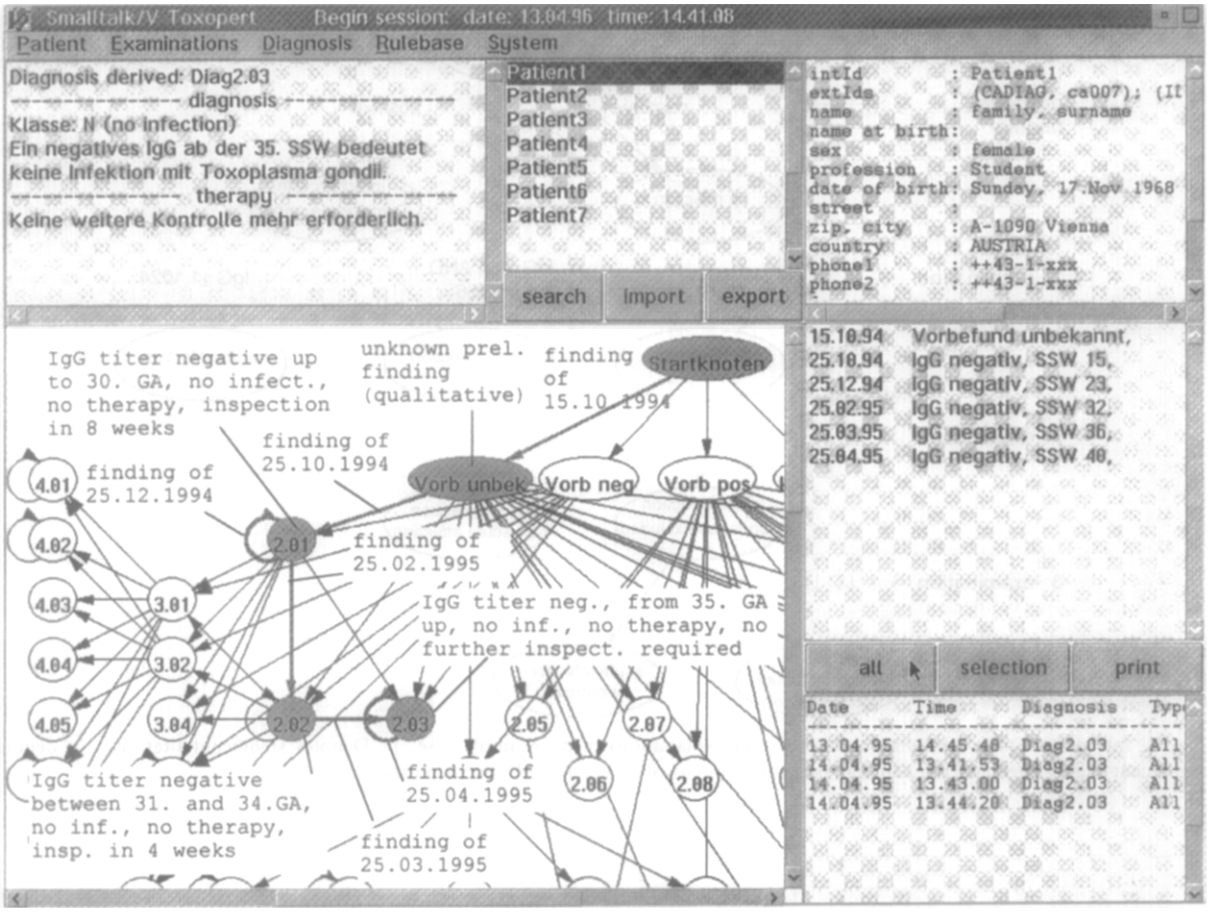

Fig. 5. Interpretation and examinations of a seronegative patient. The german abbreviation *SSW* stands for GA. The interpretive path in the subpane at the bottom left is displayed in enhanced mode. It connects the states Start, Vorb unbek (previous examination unknown), 2.01, 2.02 and 2.03 with each other. Each transition from one state to another corresponds with an examination displayed in the middle right window. Each state refers to the interpretation of all test results from the start node up to the current state. The states are coded by numbers *x.yy* whereas *x* refers to the level of the interpretive path and *yy* is a consecutive number. The last coloured state contains the inferred interpretation which is displayed in the top left window.

4. Results

4.1. TOXOPERT-I system

The upper pane in the middle of the screen contains a list of all patients. After a patient is selected (with a mouse button), the corresponding administrative patient data, e.g. date of birth, address and last name, appears on the upper right pane which is read-only. Below the patient data the corresponding examinations are displayed. The examinations appear in a multiple selection list which means that each examination can be selected or deselected to be included or excluded from the subsequent interpretive process. Below the patient's examinations three buttons are available: *all*, *selection* and *print*. *All* starts the interpretive process including all examinations of the currently selected patient. *Selection* starts the interpretive process with the selected examinations of the respective patient only. *Print* causes the printout of patient's data, examinations and the resulting interpretation. If an interpretation can be derived (no error is reported), the interpretive text (interpretation) and recommendations for therapy are displayed in the upper left window. At

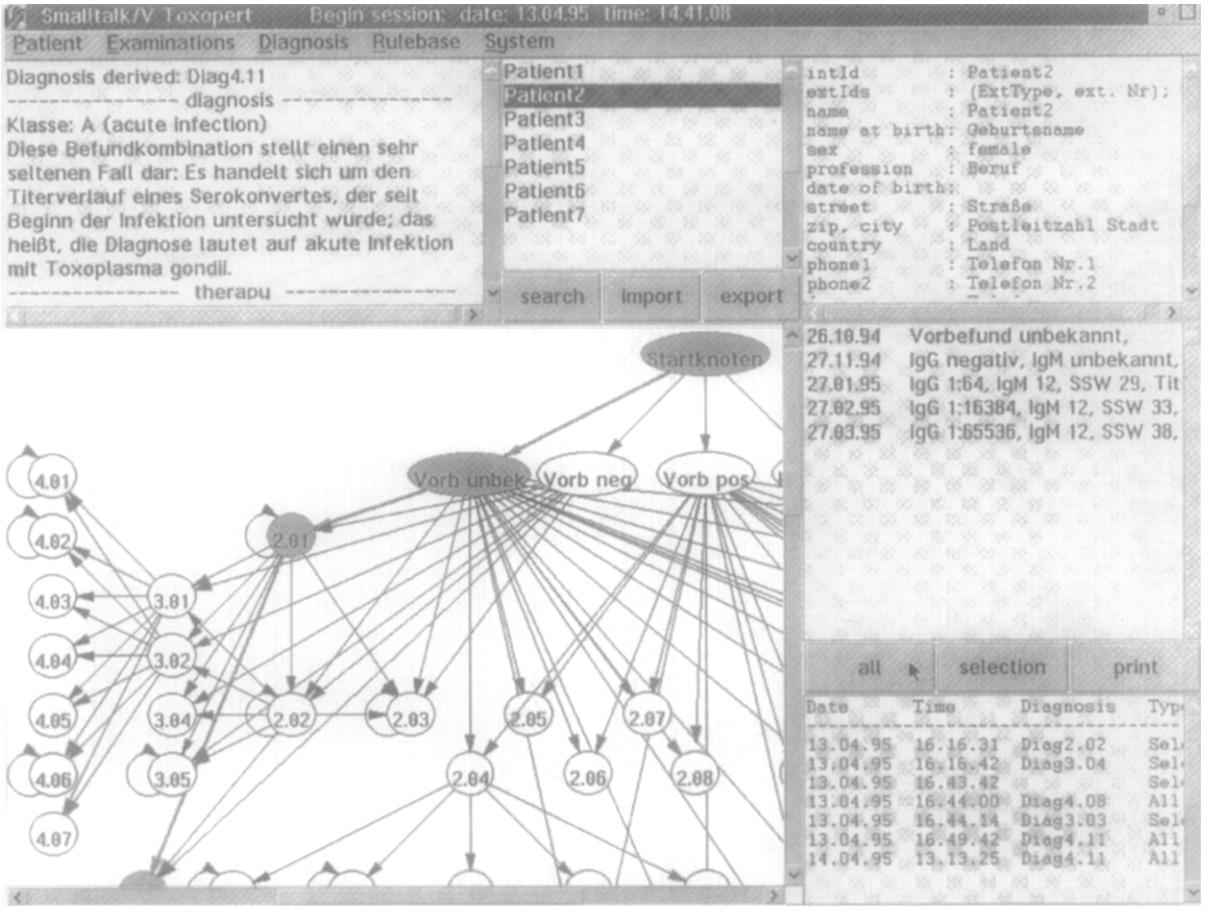

Fig. 6. Interpretation and examinations of a seroconversion. No cyclical transitions were used in this example. The interpretive path is not entirely visible in the small knowledge base subpane. Interpretation with therapy recommendations (acute infection with chemotherapy) are displayed in the top left window.

the same time the interpretive path is marked in the lower left window. (Colors and line widths are adjustable.) The interpretive process is documented in the lower right window (interpretation history). The following examples of system runs show typical interpretations.

The first example (Fig. 5) shows the interpretive path of a seronegative woman. Up to the 30th GA the interpretation and the recommendations for therapy remain the same as IgG was always negative. For this reason a cycle is used at that point. After the 35th GA no further tests are recommended and the interpretive process stops here.

Fig. 6 depicts an acute infection due to a seroconversion (i.e. a seronegative individual 'converts' to seropositivity). This constellation is a confirmation of an acute infection.

Each window can be zoomed. The example in Fig. 7 shows the decision graph with the entire interpretive path of the prior case.

Below the patient window (in the upper middle of the main window) there are three buttons: *search*, *import* and *export*. These buttons are used to organize the patients' data. *Search* looks after a certain patient, identified by name or an internal identification number. Additional external identification numbers can be entered. *Import* and *export* are functions that allow patients' data to be loaded from or saved to an ASCII file.

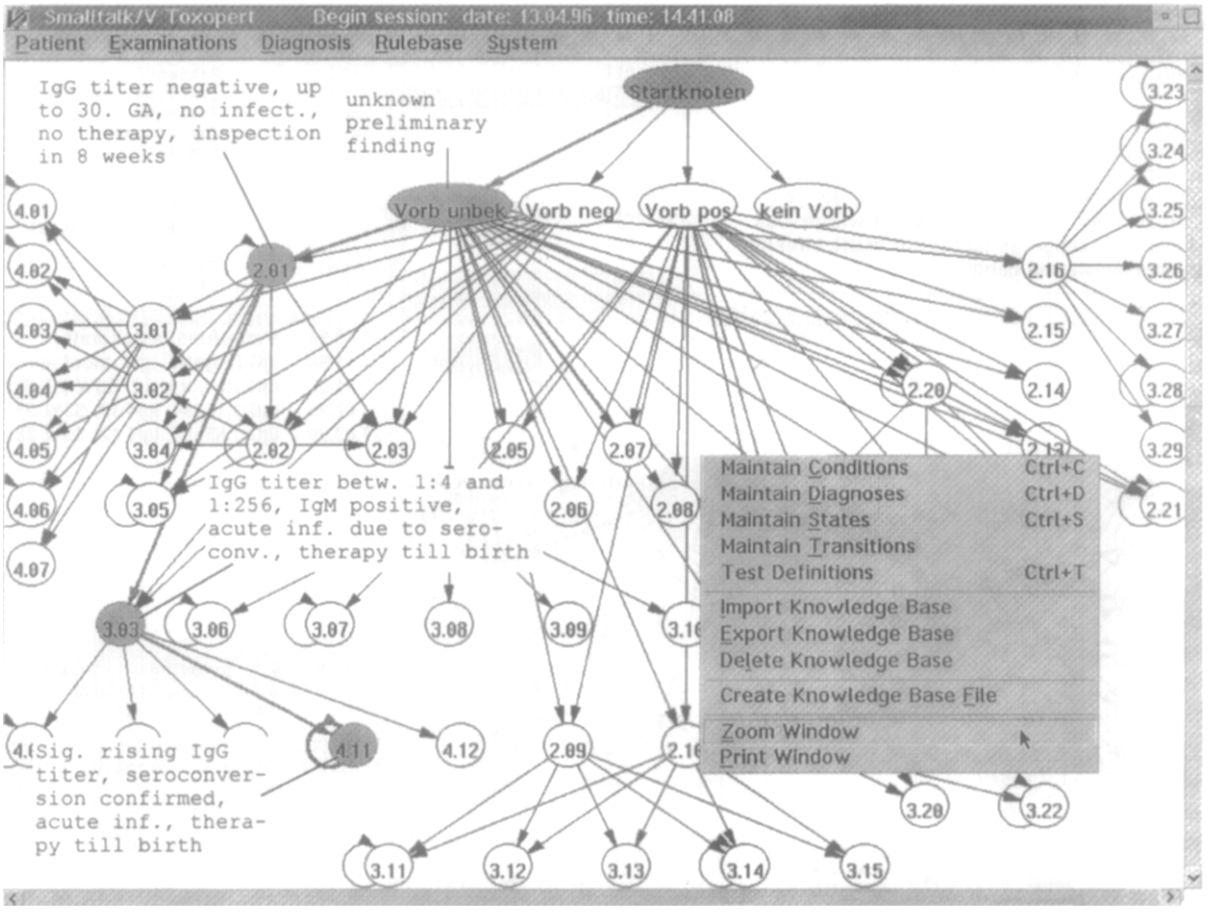

Fig. 7. This figure shows the visualized knowledge base. The states (numbered ellipses) correspond with interpretations and the transitions (arrows) correspond to examinations (collection of test results). This case shows the interpretive path (enhanced states and transitions) of a seroconversion which converted from negative IgG titer to a positive IgG titer between 1:4 and 1:256 with a following significant IgG titer rise. The interpretation (interpretation text and recommendations for therapy) of state 4.11 can be separately requested and printed with the corresponding test results and patient's data.

Any part of the knowledge base can be modified. In addition, each kind of data can be loaded from and saved to an ASCII file. In order to expedite the routine laboratory work a function called *automatic interpretation* has been implemented. It enables the laboratory worker to print out interpretive reports of all patients without starting the inference for each patient separately. A snapshot of each subpane can be sent to the printer. A snapshot of the decision graph requires the printer to support postscript format.

4.2. TOXOPERT-I shell

TOXOPERT-I consists of the knowledge base and the shell component which can support other knowledge bases as well. Due to the system requirements listed (Section 2.2) principal design considerations leading to incorporated shell-features became necessary. The creation of certain test units, e.g. IgG, IgM and GA (gestational age), defines the purpose of the expert system. Consequently, the knowledge base can be expanded to new serological tests if further laboratory tests are established.

5. Evaluation

In order to evaluate the performance of TOXOPERT-I a retrospective study group of 1000 pregnant women taken from the routinely-used database of the toxoplasmosis laboratory of the University Children's Hospital of Vienna was conducted. The serological data were taken from 562 women with at least two consecutive blood specimens, which were obtained during pregnancy at intervals greater or equal 4 weeks, and from 438 women with one blood specimen only. Routinely, for interpretation of infection with *Toxoplasma gondii*, two independent test systems are used: IgG DT and IgM ISAGA. A total of 1673 sera with a Sabin-Feldman dye test each were performed, and, in addition, 441 IgM ISAGA test results were gained.

All test results were obtained in their quantitative form: DT as a titer and ISAGA as an index number between zero and twelve. An ISAGA index from zero to five denotes a negative, from six to eight a borderline, and from nine to twelve a positive serological result. Moreover, GA was obtained as an integer where the reasonable domain ranges from one to 40.

According to the idealized titer curve of toxoplasma infection (Fig. 1 and corresponding explanation) TOXOPERT-I classified test results into one out of four interpretation groups: *seronegative*, *latent*, *acute* and *insufficient or inconsistent data*. These four interpretation groups are integrated into the knowledge base for evaluation purposes so as to help us compare TOXOPERT-I's interpretations with the clinician's interpretations. The presently available 66 different interpretations are distributed throughout the interpretation groups as follows: four *seronegative*, 18 *latent*, 19 *acute* and 25 *insufficient or inconsistent data*. An interesting side-effect is that only four interpretations are necessary to interpret *seronegative*, which represents the most frequent result, whereas 25 interpretations are required to cover all possible combinations of *insufficient or inconsistent data*, which is the least frequent result in our study. An almost equivalent amount of interpretations is necessary to differ between *acute* and *latent*. All interpretation groups are based on observations during one pregnancy:

1. *Seronegative*
   If all sera remain negative during pregnancy (first test to n-th test), there is no infection with *Toxoplasma gondii*.
2. *Latent*
   If the measured value of the first IgG test is between 1:4 and 1:256, and no significant titer rise (i.e. titer rise < two titer steps) is observed in consecutive tests, or if the IgG titer of the first test is between 1:1024 and 1:4096 and, additionally, no significant titer rise in consecutive tests is observed (without and even with IgM seropositivity; called *latent infection with IgM titer persistence*), the interpretation group is called *latent*.
3. *Acute*
   If the first test is negative and the following tests are positive (seroconversion during pregnancy, high fetal risk for infection), or if the IgG titer of the first test is positive (between 1:4 and 1:4096) and a significant IgG titer rise in the second test is observed (only with IgM seropositivity, denoted as primary high titer with titer rise) the interpretation group is called *acute*. A first test $\geq$ 1:16384 always indicates an acute infection.
4. *Insufficient or inconsistent data*
   If none of the above cases can be derived, e.g. a significant rise of DT in combination with a negative IgM ISAGA, *insufficient or inconsistent data* have to be concluded.

The clinician grouped the test results on the basis of his own knowledge of *Toxoplasma gondii* infection into the same interpretation groups mentioned above. The clinician's interpretations were used as the gold standard to be compared with TOXOPERT-I's interpretations (Table 2).

As can be noted neither an acute infection nor a seronegative case was misclassified. The seronegative cases are just as diagnostically relevant because a seronegative person is susceptible to infection at anytime and must be observed. Therefore follow-up-serology is necessary to detect infection in the ongoing pregnancy.

We can calculate a prevalence rate of antibodies against *Toxoplasma gondii* of 37% for our study (the acute and the latent infections). This is in agreement with the generally accepted prevalence rate for women between 30 and 40 [8]. In 17 of

Table 2
Performance of TOXOPERT-I

TOXOPERT-I ⇓ Clinician ⇒	Acute	Latent	Seronegative	Inconsistent or insufficient data	Total
Acute	17	0	0	0	17
Latent	0	361	0	7	368
Seronegative	0	0	606	0	606
Insufficient or inconsistent data	0	2	0	7	9
Total	17	363	606	14	1000

1000 cases (1.7%) acute infections were detected and no cases of seroconversion were observed. Table 3 shows sensitivity, specificity and accuracy rates for each interpretation group.

Fourteen cases out of 1000 (1.4%) could not be classified as *acute*, *latent* or *seronegative* (denoted as *insufficient or inconsistent data*); nine cases out of 1000 (0.9%) were misclassified.

The classes *acute* and *seronegative* achieved absolute interpretive certainty, the latent infections were correctly classified in 99.10%.

The sensitivity of the interpretation group *insufficient or inconsistent data* of 50% results from the few cases observed (seven) in that group in comparison with the seven false negative cases, which have been misclassified as latent infection.

6. Discussion

A fundamental precondition in follow-up testing is the minimum 3–4 week interval between two consecutive blood specimens before a significant titer rise can be measured. Less then 3 weeks between tests without significant rise can not exclude the possibility of a significant rise. The current knowledge base does not take into account whether there are at least 3 weeks between two examinations or not. The worst case could be an acute infection misclassified as a latent infection due to the missing significant titer rise because the time between those two examinations was too short. No such case occurred in our study. However, to avoid such a possible misclassification an improvement of the knowledge base that takes the time difference between two examinations into account is under construction.

If there had been false positive cases in the *acute* interpretation group, they would have been a result of the sharp boundaries of TOXOPERT-I's conditions. This situation could arise from a first negative finding where the IgG titer is not exactly negative but in between negative and 1:4. For the expert system, we have to compress our knowledge in a suitable way. Consequently, we have to decide between negative and 1:4. If the decision comes to negative and the IgG titer of a second finding is 1:4, TOXOPERT-I has to derive the interpretation *seroconversion* because of the IgG seroconversion. In contrast, the clinician knows that the very early discovery of an acute infection where the IgG titer is still only 1:4 is highly improbable. In addition, the first negative finding might result from inaccurate test performance in some laboratories, a facet which the experienced clinician takes into account. As a result, the clinician eventually will interpret *latent infection*. Nevertheless, TOXOPERT-I's interpretations are on the safe side. No case of acute infection was misclassified as latent or seronegative.

If at anytime during that inference no transition can be satisfied by the current examination due to an incomplete decision graph on that level then there can not be any interpretation deduced and an error is reported. We found five cases out of 1000 (0.5%), where an error was reported due to an incomplete decision graph (insufficient or inconsistent data). The clinician classified these five cases as inconsistent data as well. On the other hand, in spite of insufficient information, the clinician is able to make his most likely interpretation. This is not a real problem since a test where TOXOPERT-I reports an error message will be repeated.

Table 3
Sensitivity, specificity and accuracy rates of TOXOPERT-I

Interpretation group	Sensitivity (%)	Specificity (%)	Accuracy (%)
Acute	100.00	100.00	100.00
Latent	99.45	98.90	99.10
Seronegative	100.00	100.00	100.00
Insufficient or inconsistent data	50.00	99.80	99.10
Total	99.10	99.70	99.55

At present it is the knowledge engineer's responsibility to make sure there is a complete and consistent decision graph. In future development a consistency checker is proposed which should draw the knowledge engineer's attention to contradictions and incompleteness in the knowledge base.

In [9] an expert system for the interpretation of chemical pathology reports, called PEIRS, is proposed which does not suffer from the possible incompleteness mentioned above. In this system a path traversed by a patient is described by ripple down rules. If a rule is true the case follows the right branch; if false the case follows the left branch. The final interpretation is that of the last true rule (similar to TOXOPERT-I). However, each condition is split into a part a and a part $\neg a$ which would enlarge our decision graph substantially. The knowledge base of PEIRS was designed to be continuously improved by the user during routine work. Consequently it will never be 'complete' whereas changes in the knowledge base of TOXOPERT-I can be easily made by the user as well but are not part of daily routine. The knowledge base of TOXOPERT-I is based on the theory of toxoplasma infection in contrast to the case-based knowledge used in PEIRS.

A fuzzy approach that tries to calculate the onset of infection based on the idealized titer curve is proposed in ONSET [10,11]. In this case the idealized titer curve is implemented as a fuzzy course representing the membership of a time interval relative to the onset of infection to each possible test value. The most probable time of onset of infection can be calculated by comparing the known titer curve with the test results. Then the degree of compatibility of given test results with a fuzzy titer curve relative to possible onsets is derived. If more than one test system is performed on a patient (e.g. DT and IgM ISAGA) the method is applied to all tests and the result is a combination of all single distributions. The retrospective study presented in [10] was based on the same data as in this study. However, the studies can not be compared for the following reasons:

1. According to the premises previously mentioned the data for our study had to be adapted. This means if two blood specimens were taken immediately one after the other then they were treated as one blood specimen. Tests that were done in a different laboratory were eliminated. Furthermore, the tests had to meet certain knowledge base assumptions, e.g. in a certain situation an IgM ISAGA is necessary in order to rule out the interpretation. If an IgM ISAGA could not be found in the routinely-used database we tried to complete the data from patient records which were stored in conventional files.
2. We only distinguished between four interpretation groups instead of five in [11]. We combined the group *acute* with the group *suspected acute* and formed the interpretation group *acute* because from a therapeutic point of view, treatment is required in both cases.
3. The clinician's interpretations varied considerably according to differing test results, fewer interpretation groups, and different clinicians.

Another expert system in the matter of toxoplasma interpretation can be found in [12]. The knowledge base described in that paper uses more and various test systems alternatively, e.g. DA, SFT, IFT, IgM-ELISA, IgM-Blot, IgA-Blot,

PCR, IgE-ISAGA, in comparison with TOXOPERT-I's knowledge base. Therefore, it is harder to generate a patient's titer curve referring to the approach that we have employed, because the same test system must be applied the whole time. General screening programs should concentrate on fewer, but highly sensitive and specific serological tests. The expert system mentioned in [12] is based on Pro.M.D., an expert system shell designed for knowledge-based interpretation systems [13]. Unlike TOXOPERT-I Pro.M.D. makes use of its own description language to formulate its knowledge base.

7. Conclusion

Since the clinician's knowledge is implemented into TOXOPERT-I's knowledge base it cannot achieve better results than the clinician. The measure for the quality of the knowledge base is how close the computer-assisted interpretations are to the clinician's interpretations. The performed study does not prove how well interpretations were done but rather shows how close they compare to the clinician's interpretations. TOXOPERT-I's goal to focus on the few relevant acute infections in order to draw the clinician's attention to those cases has been attained. Additionally, the design of the current knowledge base and the connected questions led to a deeper understanding in the matter of toxoplasma infection.

The graphical user interface in combination with the visualized decision graph enhanced acceptance by the laboratory staff to test and work with the new system which was very helpful during the introductory period. It has been shown that TOXOPERT-I is able to achieve high interpretive sensitivity, specificity and thus total accuracy in the matter of interpretation of infection with *Toxoplasma gondii*.

8. Hardware and software specifications

TOXOPERT-I has been programmed in Smalltalk/V and requires a PC or compatible computer (Intel 80386 or higher). It runs under OS/2 2.1 and Smalltalk/V for OS/2 Release 2.0, it does not run under OS/2 Warp. The program (including Smalltalk/V) occupies 7.5 megabytes of hard disc storage. Eight megabytes of RAM (or more) and a high resolution monitor (e.g. 16", 768×1024) are recommended. To print the resulting decision graphs one needs a printer that supports postscript format.

Acknowledgements

The authors are grateful to Friedrich Steimann, M.Sc., Ph.D., for assisting in the implementation work and for his thoughtful comments in the preparation of this paper. This research was supported by the *Hochschuljubiläumsstiftung der Stadt Wien 1996.*

References

[1] H. Aspöck and A. Pollak, Prevention of prenatal toxoplasmosis by serological screening of pregnant women in Austria, Scand. J. Infect. Dis. Suppl. 84 (1992) 32–38.

[2] J.S. Remington and R. McLeod, Toxoplasmosis, in: International Textbook of Medicine, Vol. II, Medical Microbiology and Infectious Disease, ed. A.I. Braude, pp. 1521–1534 (W.B. Saunders, Philadelphia, 1981).

[3] P. Thulliez, Screening program for congenital toxoplasmosis in France, Scand. J. Infect. Dis. Suppl. 84 (1992) 43–45.

[4] H. Flamm, H. Aspöck, O. Picher and H. Werner, Die Toxoplasmose-Untersuchung von Schwangeren und Neugeborenen, Österreichische Ärztezeitung, 30 (1975) 15–17.

[5] G. Desmonts and J. Couvreur, Toxoplasmosis, in: Current Diagnosis, 7, ed. R.B. Conn, pp. 274–287 (W.B. Saunders, Philadelphia, 1985).

[6] O. Thalhammer, Toxoplasmose und Schwangerschaft, Zentralblatt Gynäkol. 105 (1983) 1086–1092.

[7] S. Nagy, Eine Shell für medizinische Expertensysteme in der Serologie mit einer Wissensbasis für das Toxoplasmose-Screening, Tech. Rep. No. MES-9/1993, Department of Medical Computer Sciences (University of Vienna, Austria, 1993).

[8] J.S. Remington and G. Desmonts, Toxoplasmosis, in: Infectious Diseases of the Fetus and the Unborn Infant, 3rd edn., eds. J.S. Remington and J.O. Klein, pp. 89–195 (W.B. Saunders, Philadelphia, 1990).

[9] G. Edwards, P. Compton, R. Malor, A. Srinivasan and L. Lazarus, PEIRS: a pathologist-maintained expert system for the interpretation of chemical pathology reports, Pathology 25 (1993) 27–34.

[10] F. Steimann, M. Hayde, B. Panzenböck, K.-P. Adlassnig and A. Pollak, Fuzzy support for serodiagnosis: the ONSET Program, IEEE Eng. Biol. Med. 13 (1994) 705–709.

[11] F. Steimann, M. Hayde, B. Panzenböck, K.-P. Adlassnig and A. Pollak, Computer-assisted diagnosis of Toxoplasma gondii infection during pregnancy, in: 14th Eur. Cong. of Perinatal Medicine, Abstracts, June 5th–8th, ed. Kirjapaino Tapo Oy, p. 151 (Helsinki, 1994).

[12] U. Groß, J.P. Schröder, B. Pohl and J. Heesemann, Strategien der Befundinterpretation bei der Toxoplasmose-Diagnostik—Erste Ergebnisse bei der Entwicklung eines wissensbasierten Systems, Labormedizin 18 (1994) 552–557.

[13] B. Pohl, C. Beringer, S. Walther et al., Trendelenburg, Neue Verfahren zur Erstellung wissensbasierter Befundungssysteme mit der Expertensystemschale Pro.M.D., Labormedizin 18 (1994) 577–582.