Knowledge-based System for Urinalysis

Fabrício Henrique Rodrigues, José Antônio Tesser Poloni, Cecília Dias Flores and Liane Nanci Rotta
Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil

Keywords: Ontology, Knowledge-based System, Knowledge System, Expert System, Bayesian Networks, Health Informatics, Urinalysis.

Abstract: Urinalysis is a very important test of laboratory medicine, providing valuable information about metabolism, kidney, and urinary tract. For several reasons, including lacking of professional qualification, it does not receive the proper attention, what prevents it to achieve its whole power. Considering that, a knowledge-based system for decision support in urinalysis could help to change this situation, being useful to professional training, decision support during the process or even the automation of the test. This paper proposes the development of such a system, employing ontologies, Bayesian networks, and templates of cognitive tasks to treat domain knowledge. Then, urinalysis is briefly discussed and system architecture is presented, as well as the current state of the work and future steps.

SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

Urinalysis is probably the earliest test of laboratory medicine (King, Strasinger, 2008). It can be defined as the testing of urine with procedures commonly performed in an expeditious, reliable, accurate, safe, and cost-effective manner (CLSI, 2001). Nowadays, it is an integral part of the patient examination and is composed by the following main steps:

- Urine collection and storage: obtains the sample and stores it until the analysis itself;
- Direct Observation: examines colour, turbidity, and odour of urine. It is falling out of favour in light of technological advance;
- Physicochemical analysis: carried out by means of dipstick – a plastic strip with reactive areas that gives an approximate estimation of some physical and chemical parameters of urine (e.g. density, pH, albumin). This estimation is detected through colour change of the respective reactive areas;
- Microscopy: it is done over a spot of urine in a microscope slide and identifies the particles in it (e.g. cells, crystals, microorganisms), sometimes using some auxiliary tools (e.g. polarized light, sediment stains). The same slide is analysed tens of times, in different microscopic fields (i.e. regions of the slide). After each field analysis, the observed findings are registered.

Even though inexpensive and dealing with an easily collected body fluid, urinalysis is a very important test. It can provide valuable information about many of the body's major metabolic functions, as well as the condition of the kidney and urinary tract (King, Strasinger, 2008).

However, in spite of its importance, this laboratory exam has not received the proper attention, what prevents it to achieve its whole power. One of the main expressions of this is that, generally, the urinalysis is too focused on the physicochemical analysis, leaving microscopy to a secondary role, being performed without correct methods, equipment, and professional qualification. This way, the reported results relies too much on an approximated examination of physicochemical parameters, with significant particles being missed or misinterpreted in microscopy – which means missing valuable information about the patient (Fogazzi, Verdesca, Garigali, 2008).

In order to change this scenario, (Fogazzi, Verdesca, Garigali, 2008) point out the following requirements:

i. Use of correct method for patient preparation and urine collection and handling;
ii. Capability to identify the most important particles in urine;
iii. Knowledge of clinical meaning of the urine particles;
iv. Capability to arrange urinary findings in a clinical context.

Except for (i), all the given requirements are about

Rodrigues F., Poloni J., Flores C. and Rotta L..
Knowledge-based System for Urinalysis.
DOI: 10.5220/0004952305140519
In *Proceedings of the 16th International Conference on Enterprise Information Systems* (ICEIS-2014), pages 514-519
ISBN: 978-989-758-027-7
Copyright © 2014 SCITEPRESS (Science and Technology Publications, Lda.)

pure cognitive and informational tasks, which may be suitable to computational modelling. Such requirements reveal three different facets that are necessary to take into consideration in order to rightly portrait the domain – with specific representational tools suitable for each of them.

The first facet is the representation of the complexity of the concepts involved in the task. This aspect is present in the information needed both to recognize particles in urine and define their clinical meaning as well as to describe all the findings and their clinical contexts. For that it may be useful to employ ontologies. An ontology can be defined as an explicit specification of a conceptualization (Gruber, 1993). Generally, it is represented as a set of concepts, a set of relations among these concepts, a set of attributes to describe them, and other axioms about the conceptualization.

The second facet is the uncertainty inherent to medical domain (Schwartz and Elstein, 2008). Regarding urinalysis, it is mainly due to the non-deterministic nature of the relations between findings and clinical contexts (i.e. single particles or sets of findings are not always related to the same clinical condition and may be arranged in different clinical contexts) and the usual incompleteness of information (i.e. not all findings that characterize a clinical context are always present at the same time at the same sample). Such uncertain aspect of the domain can be dealt with using Bayesian networks (BN). BNs are directed acyclic graphs in which the nodes represent domain variables and the arcs represent influences among these variables (Pearl, 1985), quantified by conditional probabilities tables.

The last facet is the reasoning processes and heuristics needed to relate findings and decide what to do to next during the sample analysis. Even though the characteristics of particles are known as well as their clinical meanings and associations, it is still needed further cognitive skills to take advantage of that knowledge (e.g. selecting a tool to identify a particle, indentify inconsistencies in the findings). With the aim of modelling so, we can use task templates (i.e. reusable combinations of model elements, that supply inferences and tasks typically used to solve problem of a particular type) (Schreiber et al, 1999). Examples of these tasks are diagnosis, prediction and monitoring.

Thus, considering:

- The importance of urinalysis;
- The requirements to be met in order to allow urinalysis to reach its whole power;
- The cognitive and informational nature of such requirements and;
- The existence of computational techniques and artefacts suitable to modelling them;

it seems to be possible and useful to develop a computational system with a representation of the domain of urinalysis able to fulfil the requirements earlier mentioned. Encompassing such capabilities, this system could be adapted and used for a variety of purposes – such as professional training, decision support and urinalysis automation.

Following this hypothesis, this paper proposes the development of a knowledge-based system – i.e. that uses artificial intelligence techniques in problem-solving processes to support human decision-making, learning, and action (Akerkar, Sajja, 2009) – for the domain of urinalysis. Due to the exposed, the core of the system is planned to be built using ontologies, BNs and template tasks, each being used for modelling the respective facet of the domain. In spite of the different possible uses for it, as a first version, the system is being conceived as a decision support tool that will be used to:

- Answer questions about the domain;
- Evaluate user's hypotheses;
- Guide the user during the exam (i.e. user provides new findings to system evaluation, which returns expert advice to the user).

The rest of the paper is organized as follows: section 2 presents urinalysis knowledge needs, section 3 presents the proposed system architecture, section 4 presents the current state of the work and next steps, and section 5 brings conclusions.

2 KNOWLEDGE NEEDS

In order to develop a system aiming to help in urinalysis, it is imperative to review its context and knowledge needs. As discussed before, executing a good urinalysis is in great extent a matter of doing a good microscopy. In this way, knowledge about the fine distinctions between types of particles in urine is unquestionably mandatory for a good urinalysis.

Yet, since urinalysis is an indirect way of assessment of patient condition, achieving this objective also depends on the knowledge of the clinical conditions that can affect the urine composition – what is important not to miss relevant particles and to correctly interpret them.

These conditions may represent rather intricate processes, albeit their influence over urine is much more limited. Then, with the purpose of avoid unnecessary complexity and still improve analysis, it is possible to use urinary profiles – i.e. combinations

SCIENCE AND TECHNOLOGY PUBLICATIONS

of urinary findings associated to clinical conditions (Fogazzi et al., 2008) – to summarize such influence. There are profiles for a variety of conditions, including nephritic/nephrotic syndromes, urinary tract infection, and hepatic disorders.

Yet, in order to perform a good urinalysis, it is not sufficient to know only the clinical and visual aspects of particles. It is also necessary to better know the whole urinalysis process. This is due to the fact that there are a number of conditions and events (some of them prior to urinalysis itself) that can influence urine contents, leading it to misrepresent the real patient condition or hindering such information. These conditions and events were extracted from literature and interviews with an expert. Some of them are listed below:

- Sample contamination (e.g. patient's lack of hygiene during collection, remains of cleaning substances in collection bottle, women's period, intentional urine diluting);
- Exposure to heat or light during storage, that may degrade some substances and particles;
- Influence of conservation methods (i.e. chemical preservatives and refrigeration)
- Sample of urine that is too much pigmented (e.g. due to some medicine patient is having) painting dipstick, masking the real color change due to chemical reaction
- Confusing particles (i.e. different particles with similar morphology and visual aspects)

Besides urinary profiles and misleading urinalysis events, it is still important to urinalysis professional some additional punctual knowledge about clinical conditions, mainly related to their expression in other laboratory exams (e.g. blood). This allows verifying the result of such exams that the patient may also have been subject of (when such possibility is available) or even directly enquiring patient's physician. Such information may be useful to evaluate some hypothesis about a clinical condition that some unusual urinary finding has pointed out – and thus be able to verify if the finding is genuine or caused by an error.

Finally, there are a lot of actions that may be taken during urinalysis (e.g. use of a specific kind of microscopy, add some substance to microscope slide, ask new sample collection, inspect patient's urinalysis history available in laboratory). Mastering the context in which each action should be taken is another requirement for a good urinalysis.

Considering the urinalysis context presented here, it was identified the following main reasoning tasks performed during urinalysis:

- Assess quality of a sample and decide whether or not ask for new collecting;
- Classify sample in an urinary profile;
- Use acquired information to guide search and interpret new findings;
- Formulate hypotheses and use additional information to confirm or refute them;
- Identify incoherencies among findings;
- Identify problems/errors during the test;
- Decide which action to carry out next;

Briefly, these are the knowledge requirements for a good urinalysis. Consequently, it should be observed in order to develop a system with a meaningful knowledge model and that provides effective guidance to the user during urinalysis. The next section presents the system architecture proposed to fulfill those requirements.

3 PROPOSED ARCHITECTURE

As it was presented, urinalysis domain involves a great number of concepts (e.g. particles, substances, profiles, tools), with a strong descriptive aspect (e.g. visual aspects) and many kinds of associations between them. Taking it into account, we decided to have an ontology as the main knowledge source for the system. It is intended to cover the domain approaching three main aspects:

- Urine representation: include all particles and substances that can appear in urine, with their relations and visual attributes, as well as all of physicochemical parameters of urine;
- Clinical information: include the urinary profiles, additional information about selected clinical conditions and other laboratory tests, and a model of patient, with its key characteristics (e.g. gender, age);
- Urinalysis process: include the representation of the dipstick and all its physicochemical tests and their relation with the respective substance or parameter, all the tools and actions the analyst can take, the events and situations that can change urine composition, and associations between findings

Bearing in mind the existence and relevance of uncertainty in health domain, the system will also be composed of BNs. But, in view of its complexity in building and maintenance (Koller, Pfeffer, 1997), instead of a large BN, the system will have specific small BNs for those portions of the domain that are more sensible to uncertainty. As much as possible, the nodes of BNs will have correspondence to concepts of the ontology, so as to guarantee further interpretation of the conclusions taken from BNs in terms of ontology axioms.

Over these knowledge models, it will be developed a system with three layers, as shown in figure 1. The first is the interaction layer. It is designed to enclose the modes of interaction between user and system (e.g. patterns for questions, evidence and hypothesis providing, system answer and guidance). The interaction will be all based on ontology concepts. Auxiliary, it will also be used a lexicon for the concepts that can be referred to by different terms and a disambiguation mechanism for terms that can be mapped to more than one concept.

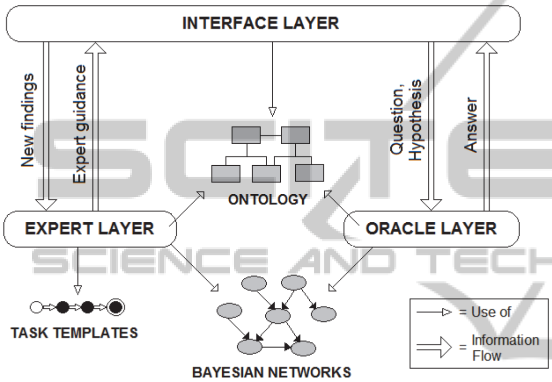

Figure 1: System architecture.

The next is the oracle layer. This layer is designed to receive questions from interaction layer in one of the predefined patterns and formulate the appropriate queries (either for ontology or some BN) to get the answer, returning it to interaction layer. Analogously, it is also designed to receive hypotheses about a sample (e.g. combinations of findings and a urinary profile user thinks that is compatible to each other), also in a predefined pattern, evaluate whether it is true or false (using ontology) or the likelihood of its truth (using some BN) and return the result to interaction layer.

Finally, there will be the expert layer. This layer is designed to simulate the expert behaviour during the exam. Thus, it is intended to be used to evaluate any new information provided by the user (which will be received through some pattern of interaction layer) and to perform the reasoning tasks enumerate in the previous section, always considering all the information already gathered during sample analysis. Aiming to ease the development of this layer, as well as to make it more understandable and maintainable, it was conceived as a chain of five task templates extracted from (Schreiber et al., 1999): monitoring, diagnosis, classification, prediction and assessment. The interactions among them are presented in figure 2.

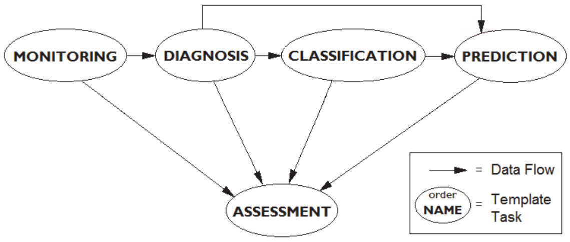

Figure 2: Task Template Chain.

Monitoring is the task of analyzing an ongoing process to verify if it behaves according to the expectations. It gets as input historical data about a monitored system and gives as output any found discrepancies from the expected values, with no further investigation of its causes. The task starts receiving new findings and evaluating its parameters against some norms. The difference is, then, classified as a type of discrepancy or as normal case.

In our system, monitoring will be used to look for inconsistencies among the findings (e.g. acid crystals found in alkaline urine), signs of false-positives and/or false-negatives (e.g. high levels of blood found on dipstick but no blood cells in microscopy) or other problems in the sample (e.g. too much epithelial cells). It will be based on ontological knowledge as well as some trigger rules learned from expert. All already gathered data will be considered. Moreover, a time index will be used to judge possible discrepancies (i.e. some discrepancies will be only considered so if discrepant values persist after the analysis of a given number of microscopic fields). As output, it will be returned any found inconsistency or sign of false-positive/false-negative or other sample problem.

Diagnosis means finding the fault that causes some malfunction in a system. The inputs of this process are symptoms and the outputs are the fault and evidences found of it. It generally uses a model of the behavior of the system being diagnosed. Diagnosis starts by taking the complaints and making hypotheses about the problem by going backwards in a causal network. Then, the actual findings are compared with the signs that should be observed for each hypothesis, excluding those in conflict with the findings. The remaining hypothesis and the observations that led to it are the output.

In the proposed system, diagnosis will be done over the output of monitoring, inferring the roots of any identified false-positive/false-negative result, problem or inconsistency. Hypothesis generation will be based on ontological knowledge and there

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

will be a BN to evaluate the likelihood of concurrent hypotheses, when more than one remains at the end. The causes of the discrepancies (or the possible ones, ordered by likelihood) will be given as output.

Classification task represents the establishment of the correct category of an object available for inspection, based on its characteristics. As input, it takes an unclassified object and gives one or more classes as output. It is done by taking candidate classes and matching their attributes with those of the object. As some attributes of the object conflicts with one of the candidate class, this is discarded. According to the matches, none, one or more than one classes can remain as the output.

The system proposed will use classification to identify the urinary profile(s) of the sample in accordance with all the data already gathered – including the output of diagnosis task. Profile definition will be ontology-based. In addition, given that it is not usual to find all the findings needed to unambiguously point to a single profile, a BN will be devised to indicate the likelihood of the alternative hypotheses. Analogously, classification task will be used to classify particles whose visual attributes are identified, but the type is not recognized by the user. In the same way, particles will be described with ontology concepts and a BN will be used to evaluate alternative hypotheses.

Prediction is the task of analyzing current system behavior to infer a description of system state in some point of the future. For that purpose, it uses a model of system behavior.

This task is planned to be used to tell what findings are likely to be seen when analyzing the next microscopic field of the microscope slide. It will use all data already gathered and the outputs of diagnosis and classification tasks. The main tool for this task will be a BN calibrated to indicate the probability of the presence of each particle in the sample. Even though it is not exactly the canonical use of the prediction, since predicting particles to be found does not represent a future state of the sample, we believe that the analogy is valid and that the general idea will be useful to our case as well.

The goal of assessment task is to find a decision category for a case, based on domain specific norms (i.e. heuristic rules). The input is data about the case and, sometimes, case-specific rules. The output is a decision category. It starts receiving a case and selecting a norm to evaluate it. The evaluation involves both case features and the available classes of decision. Depending on the result of norm evaluation, a decision class may be chosen. If it is not possible to select a decision class, another norm is evaluated. Sometimes more than one norm match is needed to assess a decision class.

We will use assessment to define what to do next in the analysis, chosen from a possible list of actions, including (but not limited to) the use of some tool (e.g. a special kind of microscope), searching for additional information (e.g. patient history) and/or a specific particle in the slide. The case representation will include all the information already gathered and the output of all other tasks. This task will be largely based in heuristics to be learned from expert. The possible actions will be described in terms of ontology concepts. As sometimes a sequence of actions may be needed, some planning routine may be run during this task.

4 STATE OF THE WORK

Several interviews with a urinalysis expert and a literature review about the domain are already made. With this material it was possible to devise the project of the system (which was briefly presented in this paper) and a plan of execution. In addition, it was already registered about 90 competency questions to guide ontology development, over 250 ontological concepts, up to 30 types of ontological relations and attributes, plus dozens of heuristics used by expert during the analysis.

Presently, we are working to formalize the concepts and to structure the ontology using the Unified Foundational Ontology (UFO) (Guizzardi, Wagner, 2005), which was chosen due to its strong logical framework and its cases of success. The ontology is going to be implemented using the Web Ontology Language version 2 (OWL2) (Grau et al, 2008). OWL2 was chosen in view of its status of W3C recommendation, which favors its stability, and the set of tools built based on it, including a powerful ontology editor – the Protégé OWL (Rubin et al, 2005). The limitations in expressiveness of OWL2 in comparison to UFO are already being considered in order to be mitigated.

Following, we are going to develop the BNs, with its structure based on the ontology model and the probabilities calibrated by the expert. The resulting BNs will be implemented using UnBBayes framework (Matsumoto et al, 2011). Next, we are going to adapt the mentioned cognitive task templates to urinalysis domain. After that, we will work on the interface layer. All software artifacts will be developed in Java.

Finally, we plan to validate our work in two ways: (i) confronting ontology and BNs individually, as well as the whole system, against real urinalysis cases, in order to verify if they are able to reach correct conclusions in comparison with expert performance and (ii) providing the system to urinalysis professionals and students and evaluate its effectiveness as training and decision support tool (i.e. whether or not it improves their capabilities).

5 CONCLUDING REMARKS

Urinalysis is a relatively inexpensive but powerful and very important laboratory test, commonly employed in patient examination. In spite of that, for numerous reasons, it has not received the needed attention to achieve its whole power, which has its roots mainly on the lack of professional qualification and insufficient knowledge about the test.

Given that most of the problem relies on cognitive tasks, suitable to computational modelling, it was formulated the hypothesis that it is possible to develop a knowledge-based system representing urinalysis domain and that this system can be useful to enhance this laboratory exam. We also believe that this usefulness can be materialized in many ways – contributing to professional qualification, decision support and even to urinalysis automation.

With the aim of testing such hypothesis, this paper proposes the development of such a system. To achieve this objective, we decided to use ontologies to model the domain, BNs to treat its uncertain aspect and task templates to formalize the reasoning tasks needed to well perform the analysis. After literature immersion and interviews with an expert in the domain, the system was designed and the ontology construction has started.

Even though being an apparently simple analysis, dealing with a single and so trivial body fluid, urinalysis revealed itself as a rather complex area. This challenging nature can be exemplified by the great amount of concepts involved (over 250, selected from an initial list of about a thousand of them), by the intense flow of information during the analysis, and the resultant intricate heuristics needed to treat it – which demanded a handful of task templates to represent. It is indeed a domain that would certainly take several years to be mastered by a novice professional.

Nevertheless, precisely due to that challenging nature of the domain, the importance of this work grows stronger. Besides an intelligent system to support urinalysis, accepting different interface layers according to the intended use, it is also being developed an ontology model that has valuable in itself. This model may serve as base for a series of useful applications for the domain, not even imagined yet. Still, considering that the methodological knowledge to be developed during this work may extrapolate its domain, our work may serve as guidance to similar initiatives in correlate domains, such as other laboratory exams.

REFERENCES

Akerkar, R., Sajja, P. 2009. *Knowledge-Based Systems*, Jones and Bartlett Publishers. Sudbury, US, 1st edition.

Clinical and Laboratory Standards Institute (CLSI), 2001. Approved Guideline GP16-A2: *Urinalysis and Collection, Transportation, and Preservation of Urine Specimens,* CLSI. Wayne, Pa., 2nd edition.

Grau, B. C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U., 2008. OWL2: The next step for OWL. *In: Web Semantics: Science, Services and Agents on the World Wide Web 6.*

Gruber, T. R., 1993. A Translation Approach to Portable Ontology Specifications. In: *Knowledge Acquisition*.

Guizzardi, G.; Wagner, G., 2005. Some applications of a unified foundational ontology in business modelling. In: *Business systems analysis with ontologies*, Idea Group.

Koller, C., D., Pfeffer, A., 1997. Object-Oriented Bayesian Networks. *In: 13th Annual Conference on Uncertainty in AI (UAI).*

Matsumoto, S., Carvalho, R.N., Ladeira, M., Costa, P.C.G, Santos, L.L., Silva, D., Onishi, M., Machado, E., 2011. UnBBayes: a Java Framework for Probabilistic Models in AI. In: *Java in Academia and Research*. iConcept Press Ltd.

Pearl, J., 1985. Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. *7th Annual Conference of the Cognitive Science Society.*

Schreiber, A. T., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., Van de Velde, W., Wielinga, B., 1999. *Knowledge engineering and management: the CommonKADS methodology*, MIT Press. Cambridge, MA.

Schwartz, A., Elstein, A. S., 2008. Clinical reasoning in medicine. In: *Clinical Reasoning in the Health Professions*. Elsevier, 3rd edition.

Rubin, D. L., Knublauch, H., Fergerson, R. W., Dameron, O., Musen, M. A., 2005. Protégé-OWL: Creating Ontology-Driven Reasoning Applications with the Web Ontology Language. In: *AMIA Annual Symposium Proceedings 2005*. American Medical Informatics Association.