UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE PORTO ALEGRE - UFCSPA
CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

Fabrício Henrique Rodrigues

Sistema Baseado em Conhecimento para Uroanálise

Porto Alegre
2015

Fabrício Henrique Rodrigues

Sistema Baseado em Conhecimento para Uroanálise

Dissertação submetida ao Programa de Pós-Graduação em Ciências da Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre como requisito para a obtenção do grau de Mestre

Orientadora: Dra. Liane Nanci Rotta
Coorientadora: Dra. Cecília Dias Flores

Porto Alegre
2015

Catalogação na Publicação

Rodrigues, Fabrício Henrique
Sistema Baseado em Conhecimento para Uroanálise / Fabrício Henrique Rodrigues. -- 2015.
100 f. : il., tab. ; 30 cm.

Dissertação (mestrado) -- Universidade Federal de Ciências da Saúde de Porto Alegre, Programa de Pós-Graduação em Ciências da Saúde, 2015.

Orientador(a): Liane Nanci Rotta ; coorientador(a): Cecília Dias Flores.

1. Ontologias. 2. Sistema Baseado em Conhecimento. 3. Informática na Saúde. 4. Uroanálise. 5. Inteligência Artificial. I. Título.

Sistema de Geração de Ficha Catalográfica da UFCSPA com os dados fornecidos pelo(a) autor(a).

AGRADECIMENTOS

Quando encerramos um ciclo como este, encerramos em débito.

Agradeço à minha família e à minha namorada, que investiram de muitas formas neste trabalho: em tempo que não pude dedicar a eles, nas minhas ebulições com que acabaram tendo de conviver e na infinidade dessas coisas da vida em que não pude me envolver apropriadamente por estar tão imerso em uma busca. Para mim foi certamente um período difícil, mas fazia sentido, pois, ao final, teria uma recompensa mensurável. Já para eles, tudo foi mais difuso – o que torna mais valioso esse apoio.

Agradeço também às minhas orientadoras, por terem viabilizado minha entrada no mestrado, pelo suporte durante esse período e pelos “trancos” quando eu precisei,

À Marta Bez, pela segurança da sua garantia de 50 anos (a que recorri um bocado de vezes),

Ao nosso grande especialista José Poloni pelo conhecimento, paciência e disposição em ajudar. De fato, é uma experiência enriquecedora conviver com alguém que realmente compreendeu sua arte,

À Mara Abel e ao Joel Carbonera pelas doses de ontologia diretamente na corrente sanguínea,

Ao pessoal da SRT e da SDSA por segurar as pontas em diversos momentos,

Ao pessoal da 4ªVT/NH, pelo pontapé inicial e pela torcida,

Ao Jérson Rodrigues, pelo apoio e pela carta branca desde o início,

À Letícia Silveira, por razões que não cabem no gibi,

E a todas as muitas (mesmo) pessoas que de alguma forma contribuiram para que este trabalho pudesse acontecer.

Às vezes fico pensando que se peso 62kg, justo seria que fossem 61kg de gratidão – e o restante de desculpas... Agradeço a Deus, por ter tanto a agradecer.

Dedico este trabalho aos meus pais, Pedro e Oraides, aos meus irmãos, Marcelo e André, e à minha namorada, Aline.

RESUMO

Introdução: A uroanálise é um importante exame laboratorial que fornece informações sobre as principais funções metabólicas, assim como sobre os rins e o trato urinário. Ela consiste na identificação de substâncias (por meio de tiras reativas) e elementos figurados (por microscopia) presentes na urina. Atualmente, por diversas razões, esse exame não tem recebido a atenção devida, o que se traduz na não-identificação, identificação errônea ou má-interpretação de seus achados, contribuindo para a confecção de um exame com qualidade duvidosa . Para colaborar na melhoria dessa situação, alguns requisitos tem sido estabelecidos. Consistindo de tarefas puramente cognitivas, esses requisitos parecem passíveis de modelagem computacional. Assim, este trabalho apresenta o desenvolvimento de um protótipo de sistema baseado em conhecimento para uroanálise com o objetivo de descrever como o conhecimento desse domínio pode ser representado e processado computacionalmente, a fim de permitir que um sistema de informação possa atuar de forma semelhante a um profissional na análise de informações sobre amostras de urina.

Métodos: O conhecimento sobre uroanálise foi elicitado da literatura e de entrevistas conduzidas com um especialista da área, tendo sido identificadas as principais atividades realizadas durante o exame (i.e. análise de tira reativa, verificação de coerência, previsão de elemenos figurados e seleção de recursos). Foi construída uma ontologia geral para o domínio assim como modelos ontológicos específicos e algoritmos para dar suporte a cada atividade específica. Tais modelos e algoritmos foram reunidos em um protótipo de sistema baseado em conhecimento para guiar o usuário durante a análise de amostras de urina, com esse reportando os achados encontrados na amostra e o sistema sugerindo elementos figurados a serem procurados, recursos a serem utilizados e incoerências entre os achados.

Resultados: O protótipo foi testado com 17 descrições de amostra de urina elaboradas pelo especialista do domínio. O intuito foi de determinar em que medida o conhecimento modelado é adequado para tratar com casos considerados importantes no cotidiano da uroanálise. Em relação à análise de tira reativa e à verificação de incoerências, a avaliação ficou parcialmente

comprometida em vista de que os exemplos utilizados não cobriram todas as possibilidades de interferências e incoerências que podem ocorrer nessas atividades. A previsão de elementos figurados atingiu precisão e abrangência de 62.08% e 79.02%, respectivamente. Considerando a seleção de recursos, o protótipo fez as mesmas escolhas prescritas pelo especialista.

Conclusões: O protótipo se comportou de forma consistente com o que era esperado dados os exemplos utilizados para teste. Além disso, a maioria das falhas apresentadas pode ser solucionadas com a adição de conhecimento no formato prescrito pelo modelo proposto, o que evidencia o seu poder e adequação na representação do conhecimento do domínio.

Palavras-chave: ontologia, sistema baseado em conhecimento, uroanálise.

ABSTRACT

Background: Urinalysis is a very important test of laboratory medicine, providing valuable information about the body's major metabolic functions, kidneys, and urinary tract. It is carried out by identifying substances (by means of dipstick) and particles (by means of microscopy) present in the urine. Nowadays, for several reasons, it does not receive the proper attention, with significant findings being missed, misidentified or misinterpreted, which calls into question the quality of the test. Some requirements were established to change this situation. Being about cognitive tasks, such requirements seem to be liable to computational representation. This way, this work presents the development of a prototype of knowledge-based system for the domain of urinalysis, with the aim of describe how urinalysis knowledge can be computationally represented and processed in order to allow an information system to act in examining a urine sample as a professional would.

Methods: Knowledge about urinalysis was elicited from literature and interviews with a domain expert, being identified the main tasks carried out during the exam (i.e. dipstick analysis, coherence assessment, particle prediction and selection of tools). An general ontology for urinalysis was constructed, as well as specific ontological constructions and algorithms to deal with the main activities involved in the test. The ontological models and algorithms were tied up to in a system that offers guidance during the examination, with user informing the observed findings and the system suggesting particles to look for, tools to use and incoherences among the findings.

Results: The prototype was confronted with 17 descriptions of urine samples elaborated by the domain expert, with the purpose of assess in which extent the modelled knowledge can deal with cases considered to be important in everyday practice. Regarding dipstick analysis and coherence assessment the evaluation was somewhat compromised, since the examples did not cover the whole extent of possibilities of incoherences and interferences that may happen in these activities. Particle prediction achieved a precision and recall of 62.08% and 79.02% respectivelly. Concerning selection of tools, the prototype made the same choices as the expert prescribed.

Conclusions: The prototype behaved consistently with what was expected given the examples it was confronted with. Most of its flaws can be overcome by addition of incremental knowledge, in the form prescribed by the proposed model – which shows its power and suitability to represent the domain.

Key words: ontology, knowledge-based system, urinalysis

LISTA DE ABREVIATURAS

API: *Application Program Interface*

HPF: High-Power-Field

KB: *Knowledge Base*

KBS: *Knowledge-Based System*

LPF: Low-Power-Field

ODP: *Ontology Design Pattern*

OWL: *Web Ontology Language*

pH: Potencial de Hidrogênio

RBC: *Red Blood Cell*

RTEC: *Renal Tubular Epithelial Cell*

SBC: Sistema Baseado em Conhecimento

SWRL: *Semantic Web Rule Language*

UFO: *Unified Foundational Ontology*

W3C: *World Wide Web Consortium*

SUMÁRIO

Capítulo I – REVISÃO DA LITERATURA

1 INTRODUÇÃO

A presente dissertação apresenta o desenvolvimento de um protótipo de sistema baseado em conhecimento para uroanálise. Tal protótipo foi construído a partir de conhecimento do domínio obtido a partir da revisão da literatura e de entrevistas com um especialista da área – conhecimento este que é brevemente retomado na seção 1.1. O domínio foi representado utilizando ontologias. Uma rápida revisão sobre esse formalismo computacional é apresentada na seção 1.2. As seções 2 e 3 apresentam a justificativa e exploram mais detalhadamente os objetivos deste trabalho. No capítulo seguinte, apresenta-se um artigo sobre o trabalho desenvolvido, com vistas a ser publicado em um periódico de informática biomédica. Finalmente, após o artigo são apresentadas as conclusões do trabalho.

1.1 Uroanálise

A uroanálise é provavelmente o mais antigo exame médico laboratorial [1]. Atualmente, a uroanálise é o terceiro maior teste de *screening* diagnóstico no laboratório clínico, precedido apenas por perfis bioquímicos séricos e pela contagem sanguínea [2a]. Ela pode ser definida como o exame de urina realizado de forma rápida, confiável, precisa, segura e com baixo custo [4], essencial para avaliar a presença, severidade e estado corrente de doenças renais e do trato urinário [5][6], bem como fornecer informações sobre as principais funções metabólicas do corpo [1][7]. A urina é composta principalmente por água, ureia, creatinina e ácido úrico [1]. Adicionalmente a essa composição básica, pode haver outras substâncias dissolvidas na urina, bem como elementos figurados (i.e. partículas microscópicas tais como células, cristais e micro-organismos) nela suspensos. São essas substâncias e estruturas adicionais que fornecem a grande maioria das informações obtidas com a uroanálise.

A uroanálise tem como etapas principais a colheita e armazenamento da amostra, a observação direta, o exame físico-químico e a microscopia [8][5], que são apresentados a seguir.

1.1.1 Colheita e Armazenagem

Procedimentos bem padronizados de colheita, transporte, preparação (i.e., fase pré-analítica) e análise da amostra urinária são a base para uma efetiva estratégia diagnóstica para uroanálise A introdução de novas tecnologias laboratoriais e automação têm melhorado a acurácia e a produtividade do processo de análise laboratorial, porém a fase pré-analítica é mandatória na obtenção da qualidade do exame, visto que dela dependem todas as demais [2].

A colheita ocorre com a micção pelo paciente em um coletor ou, em casos especiais, sondagem uretral ou punção suprapúbica [9]. A amostra é, então, armazenada até o momento de sua análise. Como a composição da urina tende a modificar-se, elementos figurados tendem a degenerarem-se e bactérias tendem a proliferarem-se caso a amostra permaneça muito tempo esperando para ser analisada [9], algum método de conservação pode ser empregado para evitar essa situação. Um desses métodos é manter a amostra a temperaturas entre + 2 a + 8°C, a fim de preservar os elementos figurados e substâncias nela contidas. Entretanto, uso de refrigeração para conservar a amostra favorece a precipitação de grande quantidade de cristais, o que pode tornar difícil seu exame (i.e. cristais podem sobrepor-se aos demais elementos figurados, dificultando sua visualização). Alternativamente, podem ser utilizados conservantes químicos para preservar a amostra, mas isso pode alterar suas propriedades químicas e a presença de elementos figurados. Dessa forma, deve ser feito um esforço para que as amostras sejam examinadas logo após sua colheita – no máximo de 2 a 4 horas após sua colheita [6].

1.1.2 Observação Direta

É a análise da amostra sem a utilização de nenhum instrumento. Nessa etapa são observados sua cor, turbidade e odor. Devido à subjetividade envolvida nesta fase e aos avanços nos métodos das fases seguintes, essa etapa encontra-se em desuso [10].

1.1.3 Exame Físico-Químico

Nessa etapa são analisadas a densidade relativa (também referida como gravidade específica) da amostra, seu pH e a presença de algumas substâncias. Atualmente é realizada utilizando-se tiras reativas (i.e. pequenas tiras plásticas, com pedaços de papel absorvente, usualmente chamados de áreas de reação, campos reativos ou almofadas reativas, contendo, cada um deles, um reagente específico para um dos parâmetros físico-químicos avaliados), em metodologia denominada “química seca”. O exame é realizado mergulhando-se a tira reativa em uma amostra de urina à temperatura ambiente. Quando em contato com a urina, cada área reativa reage com o parâmetro a que se refere, mudando de cor. Pela nova cor assumida pela pela área, obtém-se uma estimativa da intensidade do parâmetro na amostra.

Além de pH e densidade relativa, os analitos observados nessa fase são albumina (i.e. proteína cuja presença na urina pode indicar doença renal), hemoglobina (i.e. substância contida nas hemácias), esterase leucocitária (i.e. enzima contida nos leucócitos), nitritos (i.e. resultado de ação bacteriana convertendo nitrato presente na urina em nitrito), glicose (i.e. ocorre na urina quando há concentração deste elemento no sangue superior ao limiar de reabsorção renal), corpos cetônicos (i.e. resultado de jejum prolongado ou problemas no metabolismo dos carboidratos), bilirrubina (i.e. substância cujo aparecimento na urina indica provável problema hepático) e urobilinogênio (i.e. substância cuja ocorrência também pode resultar de problemas hepáticos ou de hemólise intravascular) [9][1].

A ocorrência de algumas substâncias e condições podem afetar o resultado desse exame. Urina à baixas temperaturas ou com alta densidade relativa podem reduzir a sensibilidade dos campos reativos, assim como a presença de grandes quantidades de ácido ascórbico, nitrito e outras substâncias podem causar falso-negativos para determinados analitos. Outras substâncias podem ocasionar falso-positivos, tais como alguns conservantes químicos (e.g. formalina) e componentes dos produtos utilizados na limpeza de frascos coletores (e.g. amônio quaternário, hipoclorito). As áreas reativas também podem sofrer coloração artificial por alguns corantes (e.g. piridina,

corante de ftaleína),o que ocasionar falso-positivo para algumas substâncias [1].

Essa etapa do exame é importante, pois fornece, por si só, informações valiosas sobre o paciente. Além disso, o conhecimento do resultado dos analitos dessa fase também é essencial para orientar a busca na fase de microscopia, informando sobre possíveis elementos figurados que devem (ou não) ser encontrados, bem como sobre sua morfologia e estado de preservação (e.g. urina muito densa tende a crenar hemácias, deixando-as com aspecto enrugado, ao passo que urina muito diluída e/ou alcalina tende a romper as células e destruir outros elementos figurados), auxiliando na diferenciação de elementos figurados semelhantes. Essa análise também reduz a chance de falso-positivos e falso-negativos [11] ao alertar o analista para uma microscopia mais criteriosa ou mesmo a repetição do exame físico-químico em casos de discordância entre os resultados da tira reativa e os do microscópio.

1.1.4 Microscopia

Consiste na análise microscópica do sedimento urinário. Para a obtenção do sedimento, a amostra é colocada em um tubo de ensaio de fundo cônico e centrifugada, de modo que os elementos figurados concentrem-se no fundo do tubo. Então, o sobrenadante é descartado, o sedimento é ressuspenso na pequena porção líquida ainda presente no tubo e uma pequena gota é colocada sobre uma lâmina de microscópio e coberta por uma lamínula para ser analisada [1][10]. Feita essa preparação, são avaliados 10 campos microscópicos (i.e. regiões da lâmina) de grande aumento (400x) em busca de elementos figurados que possam informar sobre as condições clínicas do paciente. Após a análise de cada campo, os elementos figurados encontrados são registrados para serem reportados ao final do exame (geralmente como uma média do número de elementos observado por campo).

Entre os principais elementos figurados que podem ser encontrados na urina estão alguns tipos de células, tais como leucócitos, hemácias em diferentes estados (e.g. isomórficas, dismórficas, crenadas, lisadas) e variantes de células epiteliais (e.g. escamosas, transicionais, tubulares renais), além de alguns tipos de micro-organismo (e.g. bactérias, leveduras). Outro tipo de

elemento figurado importante são os cilindros urinários (i.e. estruturas cilíndricas formadas no interior dos túbulos renais pela proteína de Tamm-Horsfall), que podem conter outros elementos figurados (e.g. cilindros hemáticos, cilindros leucocitários) ou aparecerem “vazios” (e.g. cilindro hialino, cilindro céreo). Também podem ser encontrados cristais (i.e. resultado da precipitação de alguns sais presentes na urina) – sendo alguns deles exclusivos de urinas ácidas e outros exclusivos de urinas alcalinas – e outros elementos figurados importantes (e.g. corpos graxos ovais, gotas de lipídio, filamentos de muco) e artefatos (e.g. material fecal, fibras sintéticas, grãos de pólen) [12].

Existem alguns tipos de elementos figurados que podem ser confundidos com outros (e.g. hemácias dismórficas e leveduras). A identificação de uma partícula vista no microscópio como um ou outro tipo confundível de elemento pode ser feita tanto considerando o contexto em que cada um dos tipos aparece (e.g. qual o valor de pH em que o elemento figurado ocorre, quais outros elementos figurados também devem aparecer) como observando características específicas cujos valores divergem entre os tipos (e.g. cor, formato, movimento).

Diferentes tipos de microscópio podem ser utilizados para o exame do sedimento (e.g. microscópio de campo claro, microscópio com filtros de contraste de fase e de luz polarizada). Também se pode recorrer a recursos adicionais, tais como corantes urinários [9] (e.g. nanquim, corante de Papanicolau) e reagentes específicos [1]. Essa variedade de recursos pode ser utilizada para destacar certos elementos figurados ou características específicas desses, o que permite sua diferenciação e classificação acertada.

Embora os procedimentos de análise manual estejam padronizados, a microscopia tradicional do sedimento urinário é trabalhosa, requer tempo de análise e é imprecisa, apresentando grande variabilidade inter-examinadores. Tem-se realizado tentativas para reduzir a variabilidade na análise manual, envolvendo o uso de amostras não centrifugadas e da automação em uroanálise. O procedimento automatizado pode minimizar o consumo de tempo e é mais factível em laboratórios com grande volume de amostras. A metodologia mais frequentemente utilizada é a citometria de fluxo ou a análise

de imagens [7], que tem relativo sucesso na contagem de alguns tipos de céluas (e.g. hemácias, leucócitos) e na indicação da presença de outros tipos de elementos figurados (e.g. cilindros) – cuja caracterização mais apurada ainda deve ser feita por microscopia tradicional.

Atualmente há várias tecnologias disponíveis para a realização do exame de urina automatizado. As tecnologias podem realizar a análise automatizada ou semi-automatizada das tiras reativas, usando a metodologia de reflectância (i.e. medem a luz refletida a partir das áreas reativas). Adicionalmente, no mercado há equipamentos que realizam a análise de urina totalmente automatizada, que homogeneiza e aspira a amostra (sem centrifugação prévia), possibilitando, com o uso de marcador, avaliar o núcleo (DNA) e citoplasma (RNA) das células e bactérias na urina nativa [13].

Todas metodologias requerem procedimentos de controle de qualidade laboratorial, pelas quais os procedimentos só podem ser realizados posteriormente à calibração dos equipamentos com soluções-padrão (i.e. com composição conhecida) a fim de garantir sensibilidade e especificidade. Isso garante a confiabilidade nos resultados apresentados, por consequência fornecido diagnóstico médico adequado e cuidado ao paciente [14].

1.1.5 Conhecimento Especialista

Além de busca na literatura, a revisão sobre uroanálise feita para este trabalho contou com entrevistas com um especialista do domínio. Especialistas são profissionais com conhecimento reconhecidamente excepcional sobre seu domínio [15], adquirido após longo período de treinamento e prática [16], o que se traduz em desempenho muito superior à média [17].

Essa capacidade notável advém da forma como organizam seu conhecimento, utilizando esquemas (i.e. estruturas abstratas que capturam regularidades de objetos e eventos, categorizando-os por seus aspectos estruturais e suas relações com outros objetos) [18]. Dessa forma, enquanto novatos tendem a classificar objetos de um domínio segundo suas percepções superficiais, especialistas o fazem em uma dimensão mais teórica [16]. Ainda, especialistas privilegiam o empacotamento do conhecimento, de forma que observações que ocorrem juntas repetidas vezes passam a assumir um significado próprio, possibilitando a automatização de processos cognitivos em

níveis mais altos, liberando recursos para inferências mais sofisticadas e aspectos novos do problema [19]. Seu desempenho diferenciado também é resultado da forma como buscam soluções, utilizando seu profundo conhecimento para reduzir os caminhos alternativos, reconhecendo e descartando aspectos irrelevantes do problema, e escolher o mais adequado. Além disso, durante esse processo, eles monitoram seu progresso, constantemente reavaliando suas escolhas.

Tais características aplicáveis à perícia de maneira geral, também podem ser observadas no caso particular da uroanálise. Enquanto grande parte da literatura dá enfoque especial aos aspectos visuais e comportamentos individuais dos elementos figurados, ao deparar-se com algum deles, o especialista não se restringe às suas percepções visuais, comparando-as com as características de todos os elementos que conhece para classificá-lo, mas considera padrões no contexto em que o elemento figurado aparece. O uso desses padrões evidencia claramente o empacotamento do conhecimento por parte do especialista.

A manifestação mais evidente disso é a organização de achados do exame no que se chama de “perfis urinários”. “Perfis urinários” são conjuntos de elementos e substâncias que são frequentemente encontrados juntos em uma mesma amostra e estão relacionados a determinadas condições clínicas do paciente [11][9]. Dessa forma, um grupo de elementos encontrados na amostra não é encarado pelo especialista apenas como achados a serem reportados, mas como participantes de um cenário maior, o que orienta a busca para seus demais participantes. Entre os perfis enumerados pelo especialista entrevistado estão o “perfil nefrítico” (caracterizado pela presença de cilindros hemáticos e/ou hemoglobínicos, hemácias dismórficas, hemoglobinúria forte e albuminúria variável), “perfil nefrótico” (caracterizado, entre outros achados, por albuminúria forte, lipidúria, cilindros e células epiteliais), “perfil hepático” (i.e. bilirrubinúria e urobilinogenúria fortes, cilindros granulosos e epiteliais, células epiteliais tubulares renais e outros) e “perfil diabético” (i.e. glicosúria e cetonúria, frequentemente acompanhadas de leveduras e bactérias).

Perfis urinários são importantes na medida de sua recorrência, pois permitem a automatização das inferências sobre quais elementos figurados devem ser esperados na amostra. Entretanto, certamente eles não abrangem todo o espectro de variações de amostras e há situações em que não são suficientes. Nesses momentos, o especialista lança mão de uma segunda forma de empacotamento, baseada em seu conhecimento aprofundado da anatomia do trato urogenital e renal, das funções metabólicas com as quais se relaciona e dos processos pelos quais partículas e substâncias formam-se, aparecem e desaparecem da urina.

Durante a análise de uma amostra, o especialista revela a existência de uma abstração desse conhecimento na forma de cadeias de eventos que ocorrem sobre certos estados da urina, causando aparecimento, alteração ou destruição de elementos, substâncias ou da própria urina, levando-a a outros estados. Assim, ao observar um achado, mesmo antes de enquadrar-se a amostra em algum perfil, o especialista navega por essa cadeia de eventos considerando o que deveria existir para que tal achado fosse possível na amostra ou o que poderia ser desencadeado por sua presença (e.g. havendo nitrito na amostra deve haver bactérias que causaram sua formação e, assim, se essas bactérias tiverem origem no interior do paciente, deve ter havido resposta imunológica, pelo que é provável que se observem leucócitos).

Esse conhecimento dos perfis urinários e dos processos que podem ocorrer na urina permite ao especialista prever quais elementos figurados podem ser encontrados na observação dos próximos campos microscópicos. Assim, tendo-se hipóteses prévias sobre a qual categoria uma partícula deve pertencer, evita-se o uso desnecessário de recursos adicionais (e.g. corantes, reagentes) para sua identificação ou diferenciação de tipos semelhantes – o que, além de trabalhoso, aumenta o tempo necessário para o exame, bem como seu custo. Além disso, reduz-se a chance de não serem percebidos certos tipos de elementos figurados por não ter sido considerada a possibilidade de sua existência.

Finalmente, como estratégia de automonitoração, a cada campo microscópico observado, em lugar de simplesmente reportar os achados, o especialista observa se esses são coerentes com o que foi observado

anteriormente no sedimento, com o resultado da tira reativa e com informações sobre o paciente. O conhecimento necessário para essa reflexão parece ser composto por padrões de inverossimilhança (i.e. situações em que os componentes e condições da urina são incoerentes), tais como a presença de cristais alcalinos em urina ácida ou espermatozóides em urina de paciente feminino.

1.2 Ontologias

A origem do termo ontologia remonta à filosofia, para a qual significa “disciplina que estuda o ser, a existência das coisas”. No contexto das Ciências da Computação, uma ontologia pode ser considerada uma especificação formal de uma conceitualização [20]. Do ponto de vista filosófico, o termo é mais próximo ao aspecto de conceitualização, referindo-se a um sistema de categorias que expressa uma visão de mundo [21]. Para efeitos de aplicação em informática, o termo aproxima-se do aspecto de especificação, geralmente referindo-se a um artefato de engenharia para representar conhecimento sobre um domínio de forma computacionalmente inteligível [22]. Por essa visão de artefato de engenharia, uma ontologia é constituída por um vocabulário específico do domínio juntamente com definições explícitas sobre o significado pretendido para os termos do vocabulário. Geralmente, sua representação é composta por um conjunto de conceitos (i.e. entidades importantes do domínio em questão), um conjunto de relações entre esses conceitos, um conjunto de atributos para descrevê-los e outros axiomas sobre a conceitualização. Adicionalmente, pode haver instâncias para reificar (i.e. criar algo concreto a partir de uma ideia) os conceitos representados.

Se representadas por uma linguagem com sintaxe e semântica formalmente definidas, ontologias permitem que motores de inferência semânticos façam inferências automáticas sobre o conhecimento nela representado. Atualmente, há várias linguagens com essa característica [23]. Uma das mais populares é a *Web Ontology Language* (*OWL*), que atualmente encontra-se na sua segunda versão (*OWL 2*), tem sido utilizada em diversas áreas, sendo adotada como padrão de fato na área de biologia e saúde, bem como tendo alcançado *status* de recomendação do *World Wide Web Consortium* (*W3C*) [3].

Em *OWL 2*, o conhecimento é representado utilizando-se indivíduos, propriedades, classes e restrições sobre propriedades. Indivíduos são instâncias dos conceitos definidos pelas classes, representando objetos específicos do mundo real. Eles podem ser agrupados a outros com características similares, formando classes, e serem descritos por atributos e relações com outros indivíduos.

Propriedades são relações binárias em nível de instância (i.e. são aplicadas apenas aos indivíduos, não às suas classes). Existem dois tipos de propriedades: propriedades de dados e de objetos. O primeiro tipo relaciona indivíduos a valores em tipos de dados simples (e.g. números inteiros, reais, texto), definindo atributos dos indivíduos (e.g. um indivíduo pode ter as propriedades “nome” com valor “Pedro” e “idade” com valor “20”). O outro tipo de propriedade funciona como relação entre indivíduos, ligando uns aos outros de forma semanticamente significativa (e.g. se os indivíduos Pedro e Maria são casados, isso pode ser representado estabelecendo-se a relação “casadoCom” entre eles).

Classes são conjuntos de indivíduos que atendem a um conceito do entendimento humano [24] e são organizados em uma hierarquia de subsunção, com classes mais específicas como subclasses de outras mais gerais, e sendo essas superclasses daquelas. Definir uma classe como subclasse de outra implica estabelecer que todo indivíduo que é membro da subclasse é também é membro da superclasse (e.g. se “Cachorro” for subclasse de “Animal”, o indivíduo “Rin-Tin-Tin” – que é um “Cachorro” – também será necessariamente um “Animal”).

Definir uma classe (ou o seu conceito subjacente) pode ser visto como definir as condições que precisam ser atendidas para que um indivíduo seja considerado uma instância da classe e outras características que esses indivíduos apresentam. Em *OWL 2*, tal definição é feita principalmente aplicando-se restrições aos tipos de relações e atributos que um indivíduo pode ter. São cinco tipos de restrições possíveis: quantificação existencial (i.e. indivíduos da classe devem estar necessariamente conectados a, pelo menos, um indivíduo/valor por meio de determinada relação/atributo), quantificação universal (i.e. indivíduos da classe podem ter como valores para determinada

relação apenas indivíduos de uma classe específica ou ter apenas valores de certo tipo ou em certo intervalo para determinado atributo), restrições de cardinalidade (i.e. indivíduos de uma classe estão conectados a um número mínimo, máximo ou exato de outros indivíduos por meio de uma relação ou têm um número mínimo, máximo ou exato de valores para certo atributo) e, por fim, propriedades podem ser restringidas de forma que seus indivíduos de uma classe relacionem-se a através de determinada relação apenas com um indivíduo específico (e.g. indivíduos da classe “Espanhol” tem como terra natal o indivíduo “Espanha”) ou tenham um valor específico para dado atributo (e.g. indivíduos da classe “Time de Futebol” tem o valor “11” para o atributo “numeroDeJogadores”) [24].

Dados esses construtos, podem ser aplicados motores de inferência sobre ontologias representadas em *OWL 2.* Entre as principais inferências que podem ser feitas estão a classificação de indivíduos (i.e. inferir que um indivíduo é instância de determinada classe dado que sua descrição atende às condições necessárias e suficientes definidas para tal classe) e subsunção de classes (i.e. inferir que uma classe “A” é subclasse de “B” quando as condições para classificar um indivíduo como “A” também permitirem que ele seja classificado como “B”).

Toda linguagem tem limitações em relação ao que pode ser representado diretamente com seus construtos. Um meio de superar algumas dessas limitações é utilizando padrões de projeto para ontologias. Para *OWL* há diversos deles, podendo ser encontrados em repositórios como [25] e [26]. Além disso, para o caso de *OWL*, é possível expressar conhecimento adicional por meio de regras na linguagem *SWRL – Semantic Web Rule Language* [27] – que são processadas pelos motores de inferência juntamente com as definições em *OWL*, de forma transparente.

1.3 Sistemas Baseados em Conhecimento

Sistemas Baseados em Conhecimento (SBCs) são programas de computador que utilizam conhecimento de um domínio para resolver problemas complexos [28]. Eles diferenciam-se de sistemas de informação convencionais por representarem o conhecimento de forma explícita (e.g. utilizando-se ontologias e regras de produção), e não embutido em seus algoritmos [29].

Assim, são compostos, basicamente, por uma base de conhecimento e um motor de inferência, o qual trabalha sobre a base de conhecimento para tirar conclusões e construir respostas aos problemas apresentados ao sistema. Geralmente, SBCs são projetados para empregar conhecimento humano para resolver problemas cuja solução requeira inteligência humana [30], caso em que também são conhecidos como “Sistemas Especialistas”.

.

2 JUSTIFICATIVA

Embora seja um exame com custo relativamente baixo e que utiliza um fluido corporal de fácil obtenção, a uroanálise é de extrema importância, na medida em que o diagnóstico de inúmeras condições nefrológicas e urológicas pode tirar proveito de uma análise bem feita. Entretanto, esse exame geralmente não recebe a atenção devida, o que o impede de alcançar todo seu potencial de auxílio ao diagnóstico.

Uma das principais expressões disso é o fato de a uroanálise frequentemente ter seu foco na análise físico-química, realizada por meio de tiras reativas. Isso deixa a microscopia em segundo plano, sendo realizada em laboratórios centrais, sem os métodos e equipamentos corretos, sem profissionais qualificados e sem considerar as informações clínicas do paciente [31]. Assim, os resultados ficam dependentes de uma estimativa aproximada das substâncias presentes na amostra, deixando-se de identificar achados do sedimento urinário ou interpretando-as de maneira errada, o que significa deixar de reportar informação valiosa sobre a condição clínica do paciente [6].

Para modificação desse panorama, Fogazzi, Verdesca e Garigali [6] apontam 4 requisitos: (i) método correto de preparação e coleta, (ii) capacidade de identificar os elementos figurados mais importantes, (iii) conhecimento do significado de tais elementos e (iv) capacidade de relacionar os achados do exame a um contexto clínico. A exceção de (i) – um processo com evidente manifestação física –, os demais requisitos são tarefas puramente cognitivas, passíveis de modelagem por técnicas de representação computacional do conhecimento – mais especificamente, utilizando ontologias para representar os conceitos do domínio e sua inter-relação. Efetuando-se tal representação, esse conhecimento poderia ser utilizado para a construção de um SBC com vistas a oferecer auxílio no cumprimento dessas tarefas.

Um SBC que modelasse esse conhecimento poderia servir como apoio na realização do exame, auxiliando na interpretação de achados e na decisão por condutas analíticas posteriores. Além disso, equipando-se o SBC com as interfaces adequadas, seu núcleo inteligente (i.e. a modelagem do conhecimento e do raciocínio empregados no processo de uroanálise) poderia ser utilizado para treinamento profissional ou outros usos educacionais (e.g.

simular um processo completo de análise, sanar dúvidas específicas em casos existentes). Ainda, a ontologia a ser construída para dar suporte ao sistema poderia servir ela própria como recurso educacional ou como referência terminológica para comunicação entre os profissionais envolvidos no processo, permitindo que ela aconteça em um mesmo contexto semântico, evitando interpretações ambíguas e outros ruídos [32].

Por fim, dadas as semelhanças entre a uroánalise e outros exames laboratoriais (e.g. análises de sangue, líquor) e a abordagem baseada em ontologias (que afastam a hipótese de uso extensivo de soluções *ad hoc*), o resultado deste trabalho poderia servir como guia para desenvolvimento de iniciativas posteriores nesses domínios correlatos.

3 OBJETIVOS

3.1 Objetivo geral

Construir um protótipo de SBC para apoio à decisão em uroanálise.

3.2 Objetivos específicos

- Elicitar o conhecimento sobre uroanálise a partir da literatura e de entrevistas com um especialista;
- Construir uma ontologia básica para o domínio de uroanálise;
- Identificar padrões e pacotes de conhecimento especialista e a forma como podem ser representados;
- Construir representações ontológicas do conhecimento necessário nas tarefas específicas do exame;
- Projetar algoritmos a serem aplicados sobre o conhecimento representado na ontologia, que articulem heurísticas utilizadas por um especialista em uroanálise no desenvolvimento de suas atividades;
- Desenvolver o protótipo de um SBC utilizando a ontologia geral e as representações ontológicas específicas como base de conhecimento e os algoritmos projetados como motor de inferência;
- Avaliar o SBC prototipado.

REFERÊNCIAS BIBLIOGRÁFICAS

1 - STRASSINGER, Susan King, DI LORENZO, Marjorie Shaub. **Urinalysis and Body Fluids.** 5 ed. Filadélfia, EUA: F. A. Davis Company, 2008.

2 - DELANGHE, J., SPEECKAERT, M. Prenalytical requirements of urinalysis. **Biochemia Medica**, v. 24, n. 1, pp. 89-104, 2014.

3 - CLINICAL AND LABORATORY STANDARDS INSTITUTE (CLSI). **Approved Guideline GP16-A2: Urinalysis and Collection, Transportation, and Preservation of Urine Specimens**. 2 ed. Wayne, EUA: CLSI, 2001.

4 - MAYO, S. et al. Clinical Laboratory Automated Urinalysis: Comparison Among Automated Microscopy, Flow Cytometry, Two Test Strips Analyzers, and Manual Microscopic Examination of the Urine Sediments. **Journal of Clinical Laboratory Analysis,** v. 22, pp. 262–270, 2008.

5 - FOGAZZI, G. B., VERDESCA, S.; GARIGALI, G. Urinalysis: Core Curriculum 2008. **American Journal of Kidney Diseases**, v. 51, n. 6, pp. 1052-1067, 2008.

6 - CHIEN, T. et al. Urine sediment examination: A comparison of automated urinalysis systems and manual microscopy. **Clinica Chimica Acta,** v. 384, n. 1-2, pp. 28–34, 2007.

7 - GRAFF, Laurine. **Análisis de Orina: Atlas Color**. Cidade do México, México: Ed. Médica Panamericana, 1987.

8 - GREENBERG, Arthur. **Primer on Kidney Diseases**. 5 ed. Filadélfia, EUA: Elsevier Health Sciences, 2009.

9 - KERR, S., MARSHALL, C., SINCLAIR, D. Emergency physicians versus laboratory technicians: are the urinalysis and microscopy results comparable? A pilot study. **The Journal of Emergency Medicine**, v. 17, n. 3, pp. 399–404, 1999.

10 - FOGAZZI, Giovanni B. **The Urinary Sediment: An Integrated View**. 3 ed. Milão, Itália: Elsevier, 2010.

11 - MUNDT, Lillian A., SHANAHAN, Kristy. **Graff's Textbook of Routine Urinalysis and Body Fluids**. 2 ed. Filadélfia, EUA: LWW (Lippincott Williams & Wilkins), 2010.

12 - ARMBRUSTER, D. A., OVERCASH, D. R., REYES, J. Clinical Chemistry Laboratory Automation in the 21st Cetury – Amat Victoria curam (Victory loves careful preparation). **The Clinical Biochemical Reviews**, v. 35, n. 3, pp. 143-53, Ago, 2014.

13 - FOGAZZI, G. B. et al. Evaluation of clinical cases in External Quality Assessment Scheme (EQAS) for the urinary sediment. **Clinical Chemistry and Laboratory Medicine**, v. 52, n. 6, pp. 1-8, 2013.

14 - AGNEW, N. M., FORD, K. M., HAYES, P. J. Expertise in Context: personally contructed, socially selected, and reality-relevant? **International Journal of Expert Systems**, v. 7, n. 1, pp. 65-88, 1994.

15 - VANLEHN, K. Problem-solving and cognitive skill acquisition. In: POSNER, M. I. **Foundations of Cognitive Science**. Cambridge: MIT Press, 1989. pp.526-579.

16 - TURBAN, Efraim. **Decision-support and expert systems: management support systems**. 3 ed. Upper Saddle River, EUA: Prentice Hall, 1993.

17 - ABEL, M., CASTILHO, J. M. V. Hybrid information systems: integrating data and knowledge management. In: CONFERENCIA INTERNACIONAL DE LA SOCIEDAD CHILENA DE CIENCIA DE LA COMPUTACION, 1993, Santiago. **Anais**... Santiago, Chile: Sociedad Chilena de Ciencia de la Computacion, 1993. pp. 137-147.

18 - ABEL, Mara. **Estudo da perícia em petrografia sedimentar e sua importância para a engenharia de conhecimento**. 2001. 239p. Tese (Doutorado em Ciências da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2001. Capítulo 4, pp. 51-65.

19 - GRUBER, T. R. A Translation Approach to Portable Ontology Specifications. **Knowledge Acquisition**, v. 5, n. 2, pp. 199-220, 1993.

20 - GUARINO, N. Formal Ontology and Information Systems. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS (FOIS'98), 1998, Trento, Itália. **Anais**… Amsterdã, Holanda: IOS Press, 1998. pp. 3-15.

21 - YANG, Y., CALMET, J. OntoBayes: An Ontology-Driven Uncertainty Model. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE FOR MODELLING, CONTROL AND AUTOMATION AND INTERNATIONAL CONFERENCE ON INTELLIGENT AGENTS, WEB TECHNOLOGIES AND

INTERNET COMMERCE (CIMCA-IAWTIC'05), 2005. **Anais**… Viena, Áustria: IEEE Computer Society, 2005. pp.457-463.

22 - MANIRAJ, V., SIVAKUMAR, R. Ontology Languages - A Review. **International Journal of Computer Theory and Engineering (IJCTE)**, v. 2, n. 6, pp. 887-891, 2010.

23 - GRAU, B. C. et al. OWL 2: The next step for OWL. **Web Semantics: Science, Services and Agents on the World Wide Web**, v. 6, n. 4, pp. 309-322, 2008.

24 - HITZLER, Pascal et al. **OWL 2 Web Ontology Language Primer**. 2 ed. W3C, 2012. Disponível em: <http://www.w3.org/TR/2012/REC-owl2-primer-20121211>. Acesso em: 28 set. 2014.

25 - ARANGUREN, Mikel Egaña et al. **Ontology Design Patterns Public Catalog**. 2009. Disponível em: <http://odps.sourceforge.net/odp/html>. Acesso em: 28 Set 2014.

26 - GANGEMI, Aldo, PRESUTTI, Valentina. **Ontology Design Patterns**, 2014. Disponível em: <http://ontologydesignpatterns.org>. Acesso em: 28 Set 2014.

27 - HORROCKS, Ian et al. **SWRL: A Semantic Web Rule Language Combining OWL and RuleML**. W3C, 2004. Disponível em: <http://www.w3.org/Submission/SWRL>. Acesso em 28 Set. 2014.

28 - HUMPHREY, S. M. Knowledge-based systems for indexing. In: FIDEL, R. et al (Eds.). **Challenges in indexing electronic text and images**. Medford, EUA: American Society for Information Science by Learned Information, 1994. pp. 161-175.

29 - JOO, W., CHA, S. S. Knowledge-Based Hybrid Expert System for Automated Interferometric Data Reduction. **Optics Lasers in Engineering**, v. 24, n. 1, pp. 57-75, 1996.

30 - HAYES-ROTH, Frederick. Knowledge systems: An introduction. **Library Hi Tech**, v. 10, n. 1-2, pp. 15-29, 1992.

31 - FOGAZZI, G. B., GRIGNANI, S. Urinary microscopic analysis: an art abandoned by nephrologists? **Nephrology Dialysis Transplantation**, v. 13, n. 10, pp. 2485-2487, 1998.

32 - GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. **International Journal of Human and Computer Studies**, v. 43, n. 5-6, pp. 625-640, 1995.

Capítulo II – ARTIGO CIENTÍFICO

Modelling Urinalysis in a Knowledge-Based System[#]

Fabrício Henrique Rodrigues (1*), José Antonio Tesser Poloni(1,2),
Cecília Dias Flores (1) and Liane Nanci Rotta(1)

1 Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre – Brazil.

2 Central Laboratory of Clinical Analysis – Irmandade da Santa Casa de Misericórdia de Porto Alegre, Porto Alegre – Brazil.

Artigo a ser submetido à revista BMC Bioinformatics

* Corresponding author: Fabrício Henrique Rodrigues, Sarmento Leite, 245, CEP: 90.050-170, Porto Alegre/RS, Brasil, e-mail: titofhr@gmail.com Phone: +55 51 3303 8836

Abstract

Background: Urinalysis is a very important test of laboratory medicine, providing valuable information about the body's major metabolic functions, kidneys, and urinary tract. It is carried out by identifying substances (by means of dipstick) and particles (by means of microscopy) present in the urine. Nowadays, for several reasons, it does not receive the proper attention, with significant findings being missed, misidentified or misinterpreted, which calls into question the quality of the test. Some requirements were established to change this situation. Being about cognitive tasks, such requirements seem to be liable to computational representation. This way, this work presents the development of a prototype of knowledge-based system for the domain of urinalysis, with the aim of demonstrate how urinalysis knowledge can be computationally represented and processed in order to allow an information system to act in examining a urine sample as a professional would.

Methods: Knowledge about urinalysis was elicited from literature and interviews with a domain expert, being identified the main tasks carried out during the exam (i.e. dipstick analysis, coherence assessment, particle prediction and selection of tools). An general ontology for urinalysis was constructed, as well as specific ontological constructions and algorithms to deal with the main activities involved in the test. The ontological models and algorithms were tied up to in a system that offers guidance during the examination, with user informing the observed findings and the system suggesting particles to look for, tools to use and incoherences among the findings.

Results: The prototype was confronted with 17 descriptions of urine samples elaborated by the domain expert, with the purpose of assess in which extent the modelled knowledge can deal with cases considered to be important in everyday practice. Regarding dipstick analysis and coherence assessment the evaluation was somewhat compromised, since the examples did not cover the whole extent of possibilities of incoherences and interferences that may happen in these activities. Particle prediction achieved a precision and recall of

62.08% and 79.02% respectivelly. Concerning selection of tools, the prototype made the same choices as the expert prescribed.

Conclusions: The prototype behaved consistently with what was expected given the examples it was confronted with. Most of its flaws can be overcome by addition of incremental knowledge, in the form prescribed by the proposed model – which demonstrates its power and suitability to represent the domain.

Key words: ontology, ontologies, knowledge-based system, knowledge system, expert system, health informatics, biomedical informatics, urinalysis

1 Background

Urinalysis is probably the earliest test of laboratory medicine [1]. It can be defined as the testing of urine with procedures commonly performed in an expeditious, reliable, accurate, safe, and cost-effective manner [2]. Nowadays, it is an integral part of the patient examination that occurs by physicochemical analysis and microscopy over a urine sample previously collected. The following description of such steps is based on [1] and [3].

Physicochemical analysis aims to identify the state of the sample concerning a variety of physical and chemical parameters. Nowadays, it is carried out by means of dipstick – a plastic strip with reactive areas that gives an approximate estimation about the specific gravity and pH of the sample, as well as the presence of some chemical substances – usually albumin, hemoglobin, leukocyte esterase, nitrites, glucose, ketones, bilirrubin and urobilinogen. The estimation of such parameters is observed through the color change of the respective reactive areas, with different colors for different levels of the parameter. Urine in low temperatures or with too high specific gravity as well as the presence of great amounts some substances (e.g. nitrites, ascorbic acid), decrease the sensitivity of some reaction areas, what may cause false-negative results for some substances. Other substances (e.g. formalin, quaternary ammonium, hypochlorite) may react with certain reactive areas causing false-positive results. Finally, reactive areas are also liable to artificial coloring (e.g. by pyridine, phthalein stain), that may cause false-positive for some substances .

Microscopy is carried out in order to identify urinary particles in the urine sediment that can inform about the patient's clinical condition. The main urinary particles are cells – e.g. red blood cells (RBCs), leukocytes, renal tubular epithelial cells (RTECs) –, crystals, microorganisms (e.g. bacteria, yeasts), casts (i.e. cylindrical structures formed inside renal tubules, that may contain other particles inside them), some other important particles (e.g. oval fat bodies, lipid drops) and artifacts (e.g. fecal material, synthetic fibers, pollen grains). Some particle types are likely to be confused with others (e.g. dismorphic RBCs and yeast cells) – being differentiated by specific features and the contexts in which they appear. Different kinds of microscope can be used in sediment

examination (e.g. bright field, phase contrast, polarized light) as well as some additional resources (e.g. sediment stains, reagents), for the purpose of identifying specific aspects of the particles or differentiate confounding ones. The same slide is analysed by observing ten different microscopic fields (i.e. regions of the slide), in magnification of 400x. The observed findings are registered to be reported as the mean number of structures observed per field.

Even though inexpensive and dealing with an easily collected body fluid, urinalysis is a very important test, providing valuable information about many of the body's major metabolic functions, as well as the condition of the kidneys and urinary tract [1]. 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 dipstick analysis, leaving microscopy to a secondary role, being performed without correct methods, equipment, and qualified professionals and disregarding the information about the patient [16]. 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 [17].

In order to change this scenario, Fogazzi, Verdesca and Garigali [17] 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 pure cognitive and informational tasks, which may be suitable to computational representation. This way, it seems to be possible and useful to develop a computational system with a representation of the domain of urinalysis able to fulfill those requirements. A system with such characteristics could be classified as a

knowledge-based system (KBSs) – a computer program that use knowledge about a domain to solve complex problems [13].

A KBS differs from traditional information systems due to the fact that it has its knowledge explicit represented (e.g. using ontologies and production rules) instead of having it embedded in its algorithms [14]. This way, it is basically composed by a knowledge base (KB) and a reasoner to make inferences over the KB and provide the answers expected from the system. Generally, KBSs are designed to employ human knowledge to solve problems whose solution is recognized to need human intelligence to be addressed [15]. In theses cases, they are also known as expert systems.

Considering the complexity and richness of the concepts of the domain – e.g. those needed to represent urinary findings, their relationships and behavior in presence of some tool – and the way its knowledge seems to be employed during the execution of urinalysis, it seems a good approach to represent it by means of ontologies. In informatics, the term ontology can be defined as an explicit specification of a conceptualization [4], generally referring to an engineering artefact to represent knowledge about a domain in a computationally intelligible way [6]. It is constituted by a specific vocabulary plus a set of explicit assumptions regarding the intended meaning of the vocabulary terms. Generally, it is represented as a set of concepts, a set of relationships among these concepts, a set of attributes to describe them, and other axioms about the conceptualization. Additionaly, it may have instances reifying its concepts.

If an ontology is represented in a language with formally defined semantics, it is possible to use semantic reasoners to automatically make inferences over its knowledge. Presently, there are several ontology languages with such feature [7]. One of the most popular is the Web Ontology Language (OWL) – currently in its second version (OWL 2) –, which has been widely used in diverse areas, even being adopted as a *de facto* standard in the life sciences community as well as achieving the status of a W3C recommendation [8]. Finally, it is possible to use semantic reasoners to make some basic inferences (e.g. classification of instances, class subsumption) over ontologies represented in OWL 2.

Following this hypothesis (i.e. it is possible to computationally represent urinalysis knowledge – specially concerning the requirements pointed in [17] – and develop a system from it), this paper presents the development of a prototype of a KBS for the domain of urinalysis, with its KB represented by means of ontologies and reasoner implemented as a set of specific algorithms to make inferences over the KB in order to support the main activities carried out during urinalysis. The aim of such effort is demonstrating how urinalysis knowledge could be computationally represented and processed in order to allow a system to act in examining a urine sample as a professional would. As a first version, the prototype was conceived as a decision support tool that can be used to guide the user (i.e. somebody that is performing the analysis of a sample) during dipstick and microscopy steps of the analysis (i.e. user provides each new finding to system evaluation, which return expert advice to the user).

The rest of the paper is organized as follows: section 2 presents methodological aspects of the work, including results of interviews with urinalysis experts, the ontological models built to represent urinalysis knowledge and the algorithms designed upon them; section 3 presents the evaluation of the proposed system by confronting it with fictitious examples of urine sample prepared by the expert, the results of the evaluation and further discussion about them; finally, section 4 brings conclusions and future work.

2 Methods

For the purpose of developing the proposal of this work we turn to urinalysis literature to elicit knowledge about the domain, mainly concerning concepts and their description and relationships that were used to build up the ontologies that make up the KB of the system. Yet, we also seek for expert knowledge to increase the effectiveness of the representations. For this purpose, we conducted a series of interviews with an urinalysis expert in order to get insights on how to efficiently arrange the knowledge of the domain, what are the main portions to focus on and how to deal with it during the test. The interviews included structured and unstructured approaches, as well as the observation of his procedures while analyzing some specimens.

Experts are professionals recognized by their great knowledge in a domain [18] and that have performance highly above average [19]. Much of

their differentiated capacity is due to organizing their knowledge in schemes – i.e. abstract structures that portrait the invariance of objects and events concerning structural aspects and relationships with other entities [20]. This way, while novices would classify objects using superficial perceptual features, experts do so in a much more theoretical dimension [21] [22]. Moreover, experts tend to rely on knowledge packages, in a way that *stimuli* that recurrently happen together are seen as a unit, having its own meaning. This allows automation of some cognitive processes, freeing up resources to more sophisticated inferences [23]. Finally, their superior performance is also due to the way they search for solutions, always using their knowledge for pruning alternative paths, discarding irrelevant aspects of the problem and choosing the best current alternative, while constantly monitoring their progress – reassessing their choices.

During the interviews, it was noticeable that these features of expertise in a broad sense are also identifiable in urinalisys case. While urinalysis literature greatly focuses on visual aspects of individual particles, the expert, when observing a particle, does not get stuck to it. Indeed, instead of relying solely on particle appearance (i.e. mentally comparing it to all other known particle types), the expert considers patterns of the context in which the particle appears. This makes evident his knowledge packaging.

The most clear sign of such packaging is the practice of organizing the findings in urinary profiles – i.e. groups of particles and substances usually found in a same specimen, which are related to some patient's clinical condition [3]. Nephritic, nephrotic, liver disorders and imunossupression are among the profiles listed by the expert. Thereby, for him, a group of findings is not just information that must be reported, but components of a bigger picture that should guide the search towards other ones.

Urinary profiles are rather important, in the extent of their recurrence, since it automates the inference about what else is expected to be found in the specimen (i.e. being found some of the findings of a profile, the others may reasonably be expected, disregarding why/how they have appeared in urine). Still, they do not cover all possible variations of specimens and sometimes the expert can not avoid entering further details about urine formation and behavior.

However, even in these situaions, he makes use of (a second form of) knowledge packaging.

The expert has great knowledge about the dynamics that underlies urine production, in pretty high level of detail. In order to efficiently use such knowledge when examining a sample, the expert seems to abstract it in a chain of events that may occur to the urine. They happen transforming urine from one state to another, covering either *in vivo* or *ex vivo* cases. Such transitions may be caused either by the appearance or destruction of substances and particles (e.g. cast formation, bilirrubin destruction by light exposure), or by changes in substances, particles or in the urine itself (e.g. urine alkalinization, cell lysis). Therefore, when observing a new finding, the expert browses such event chain, considering what would have to exist in the specimen in order to make such finding possible (e.g. if hemoglobin is found, there should be some RBCs to provide such hemoglobin) or what could be triggered by the presence of such finding (e.g. if bacteria is found, probably patient's immune response will send some leukocytes after them).

With such cognitive schema (i.e. urinary profiles and events chain), the expert can foresee the particles that are likely to be found when observing the next microscopic fields of an slide. Thus, having previous hypotheses about the types of particles to be seen avoids the use of unnecessary additional resources (e.g. stains, reagents) to identify particles or differentiate confounding ones – what is time consuming, as well as increases the cost of the test. Beyond that, expecting probable particles reduces the likelihood of overlooking or mistaking them by other types – what would mean losing important information about patient's condition.

Finally, it is possible to notice some automonitoring. After each microscopic field is observed, instead of simply register the findings, the expert checks whether they are coherent with what was previously found during microscopy, with the results of dipstick and with patient's known information (e.g. gender, clinical condition). Doing so, the expert can also assess the quality of the sample, deciding whether or not to ask for new collection.

From the knowledge gained by reviewing the literature and interviewing the expert, it was possible to identify 5 main activities performed by the analyst during the analysis of a sample:

- Dipstick analysis
- Coherence assessment
- Profile-based content prediction
- Event-based content prediction
- Selection of tools

This way, the prototype was developed to cover these 5 activities. In order to compose its KB, it was built a general ontology for urinalysis comprising its main concepts. Along with that, it was built specific ontological models to deal with each of the covered activities. As its reasoner, the system was also composed by algorithms to be run over the KB and make the inferences needed by each of the activities.

All ontological portion of the system was designed according to UFO prescriptions [24] and represented using OWL 2 – with some SWRL [12] insertions when some knowledge had to be represented as production rules. The algorithms were developed using the OWL API [25] and use the HermiT semantic reasoner [26] to make basic inferences over the ontology (e.g. classification of indiviuals). The following sections present the general ontology as well as the specific ontological models and the corresponding algorithms for each main activity.

2.1 Ontology for Urinalysis

Although each of the refered tasks has its own specific knowledge needs, there is a common core of concepts that may be useful throughout the whole process of urinalysis. Provided that, it was built a general ontology covering the following main aspects:

- Urine description: composition, some features (e.g. pH, specific gravity, temperature) and their occurrences as in vivo or ex vivo specimens.
- Findings: particles and substances that can appear in urine, their features (e.g. composition, visual aspects) and relationships;

- Patient information: a model of patient (e.g. gender, age) and some selected clinical conditions s/he may present;
- Urinalysis process: reaction areas, tools for microscopy.

In order to arrange the concepts involved in these aspects, the ontology was built upon a structure of root conceps inspired in the structure of UFO, adapting it to the domain requirements. According to this structure (Figure II.1), there are two kinds of thing: the occurrents (i.e. things that extend in time, being composed of temporal parts, such as events) and the continuants (i.e. things that persist in time, being wholly present in each moment, such as objects). Occurrents are existentially dependent on continuants, since an occurrent can solely be manifested through changes in continuants or actions they perform. For this reason, every event must have at least one continuant as participant and every situation must have at least a continuant present. In turn, continuants may be either existentially independent (i.e. objects) or dependent on other continuants (i.e. features). For this reason, objects may or may not be characterized by a feature, but every feature must characterize one object, since its existence depends on this relationship.

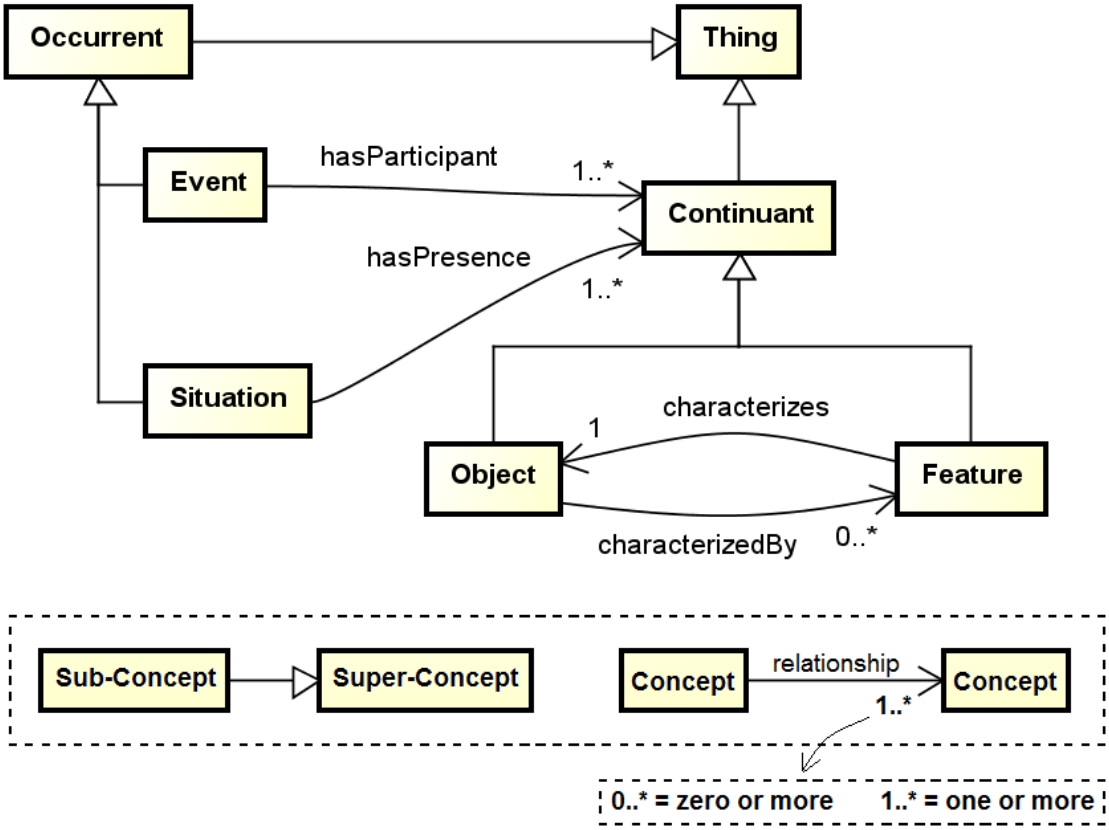


Figure II.1 - Top Ontological Model

Objects may be further classified as functional complexes (i.e. wholes whose parts has specific functions, such as cells, with their membrane and cytoplasm), collectives (i.e. sets of identical members, such as tubular fragments, which are composed by RTECs) or quantities (i.e. uncountable entities, such as hemoglobin). Figure II.2 shows a little extract of the objects in ontology. Besides the classification of some urinalysis concepts in those three types of object, it is noteworthy that there are other subconcepts of “Object” that represent categories or roles the objects take in urinalysis. This way, in addition to being functional complexes, RTECs, RBCs and pollen grains are all urinary particles when observed in a sample, although RTECs and RBCs are true components of urine, while pollen grains are not. Both nankin and hemoglobin play the role of colorings in urinalysis, but nankin is also a tool for the exam – just like microscope.

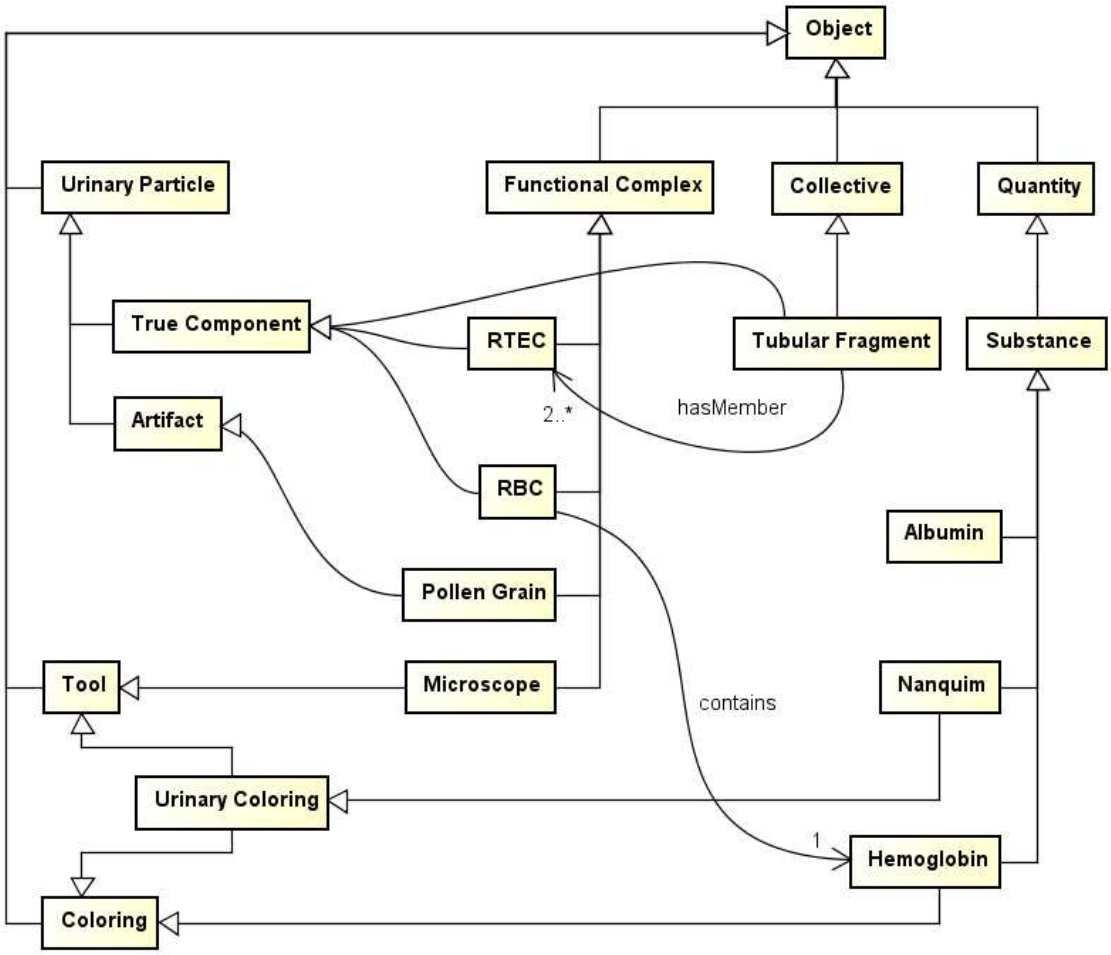


Figure II.2 - Ontological Model for Objects

Features may be either qualities or modes. As defined in [24] qualities are features expressed in terms of a point in a space of possible values (e.g. age) while modes represent conditions of something (e.g. fever) that may even be further characterized by qualities (e.g. high intensity fever). Figure II.3 brings an excerpt of the features represented in the ontology. There are several qualities, used to describe many kinds of objects (e.g. particles have colors, urine has pH, person has gender). Modes are much less numbered, being exemplified just by clinical conditions a person may present and urinary profiles, that will be discussed in a later section.

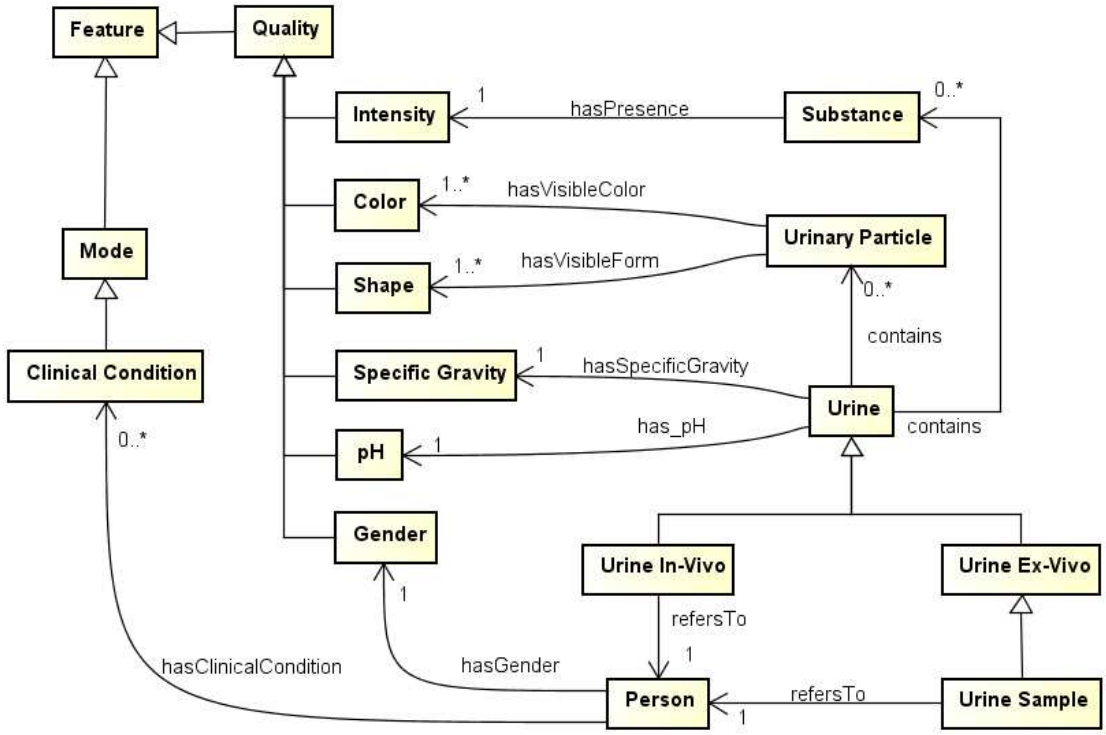


Figure II.3 - Ontological Model for Features

It is important to notice that, along with the concepts involved in urinalysis, it was also modeled relationships to suitably link their instances, allowing more complex representations of their meanings. For example, in Figure II.2, tubular fragments are defined as collectives that have 2 or more RTECs as their member. In Figure II.3 it is established that a urine sample refers to only one person. Beyond that, it was also created specific relationships to attribute features to the objects (e.g. hasGender, has_pH).

The constructed ontology is made up of over 200 concepts and a dozen relationships. All the following specific models were based on the knowledge represented in this ontology, adding new concepts and relationships as they were needed.

2.2 Dipstick Analysis

Even though it is clearly in microscopy phase that most of cognitive effort is made, great part of the knowledge that guides it is obtained in dipstick analysis, which require a careful treatment. Moreover, this task has its own dynamics that requires a different approach from the analyst, and so deserves a specific model for itself.

At first glance, dipstick analysis looks pretty simple: if a reaction area is activated, its analyte is present in the sample; if not, there is no such substance. However, taking this correlation for granted may lead to some mistakes, since in certain situations results of dipstick can diverge from urine reality. This way, understanding its ontological nature may help to avoid such problem.

In fact, a reaction area can be viewed as a sensor designed to detect presence of a substance in urine. Being a real – rather than ideal – artifact, it necessarily has imperfections that sometimes may lead to failures in its function, whose expressions are false-positives or false-negatives. False-positives may happen in two ways: by the presence of a confounding signal (i.e. that behaves as the true one and is detected as such) or by positive masking the sensor (i.e. placing something between the sensor and its observer that mimics the behavior of the sensor when activated). Similarly, false-negatives may also happen in two ways: by the presence of an inhibitor (i.e. something placed between the sensor and the sensored object, preventing its signal to be detected or acting on the sensor causing its malfunction) or by negative masking (i.e. something between sensor and observer that prevents its positive report to be perceived).

Transposing this general description to the case of reaction areas, we have the model depicted in Figure II.4. There we can see that each reaction area has a color and one analyte as its objective of detection. It can be observed as apparently activated, when it may be really the case or it may be a positive masking. If really activated, there will certainly be at least one reagent substance in contact with the reaction area and it may be either the real analyte or a confounding substance. If not activated, the reaction area is positively masked by a coloring that resembles its color when activated.

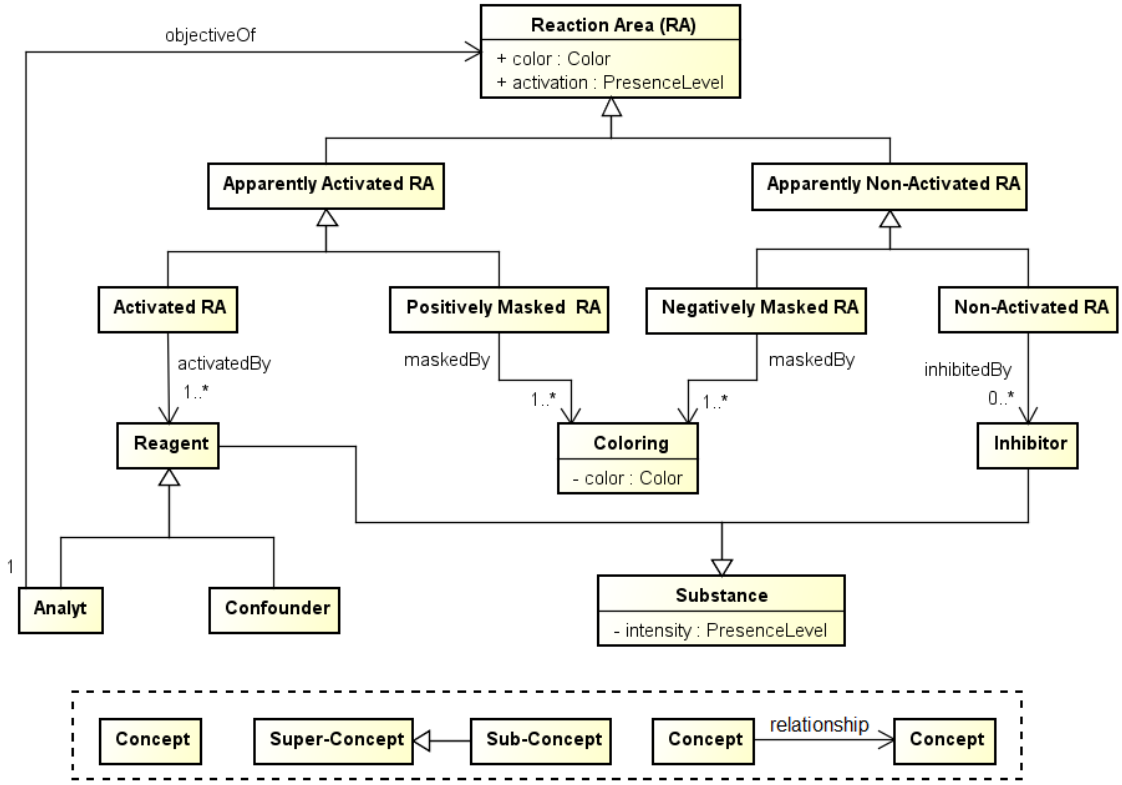


Figure II.4 - Ontological Model for Reaction Areas

Alternatively, reaction area may looks like it is non-activated. Again, it may be either really non-activated or just negative masked. If really non-activated, it is possible to have nothing to be detected or that some inhibitor is preventing the detection. Otherwise, it could be the case that some coloring that resembles the color a negative reaction area is negatively masking it – what is included just for theoretical completeness, since there is no known substance with such behavior and that can appear in urine and stick in reaction areas.

Each concept in this model was represented as an OWL class in the ontology, except for apparently activated or non-activated reaction areas and their masked cases, whose concepts were represented, for practical reasons, directly on the algorithm designed to run over this portion of the ontology. Each reaction area and its related concepts were represented according to this model (Figure II.5 presents an example for hemoglobin reaction area). Classes for specific reagents, confounders and inhibitors were created (e.g. “Reagent for Hemoglobin”, “Confounder for Hemoglobin”), with classes for the specific substances that play such roles (e.g. “Hemoglobin”, “Hypochlorite”) placed under them. All specific classes were made subclasses of their general concept

(e.g. “Hemoglobin RA” is subclass of “Reaction Area”, “Inhibitor for Hemoglobin” is subclass of “Inhibtor”). The interfering substances (i.e. confounders and inhibitors) currently modelled in the prototype are shown in Table 1.

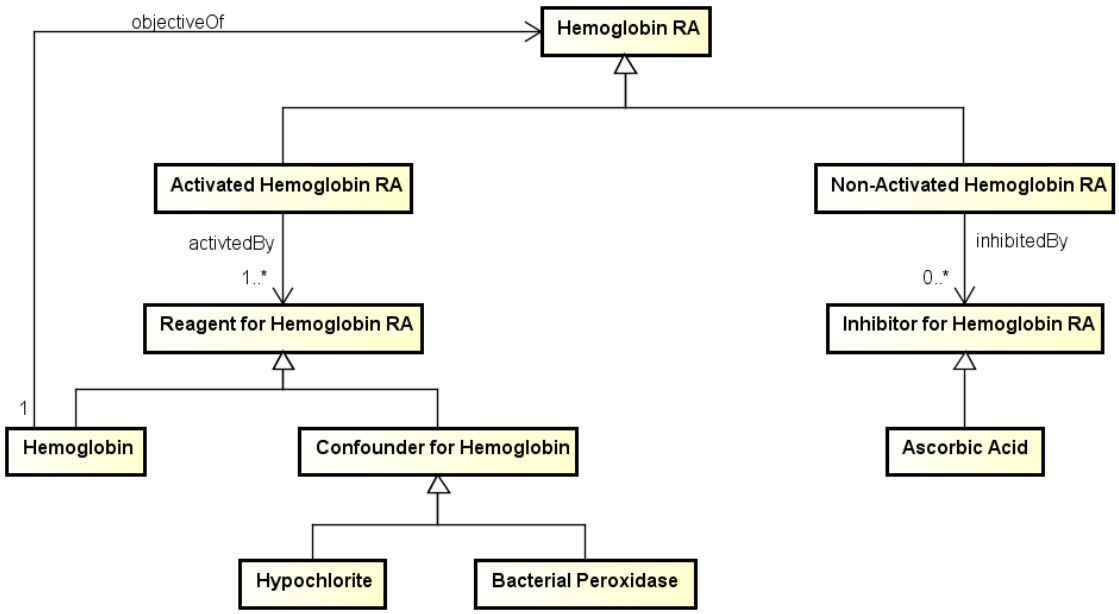


Figure II.5 - Ontological Model for Hemoglobin Reaction Area

Table 1 - Currently modelled interfering substances.

Substance	**Confounder for**	**Inhibitor for**
Hypochlorite	Hemoglobin, Leukocyte Esterase, Glucose	-
Quaternary ammonium	Albumin	-
Ascorbic Acid	-	Hemoglobin, Leukocyte Esterase, Glucose, Ntrite, Bilirrubin
Formalin	Leukocyte Esterase	Hemoglobin, Urobilinogen
Mioglobin	Hemoglobin	-
Bacterial Peroxidase	Hemoglobin	-
Urine in low temperaure	-	All analytes

To complete the representation, it was created a SWRL rule to represent that everything urine contains is in contact with everything else present in urine (i.e. if A is in urine and urine contains B, A is in contact with B). The rule is as follows

```
(1)     contains(?a, ?b), contains(?a, ?c) ->
        contact(?b, ?c), contact(?c, ?b).
```

Besides that, it was created an “Unfeasible” class and made two additional rules in SWRL for each type of reaction area: one states that if some activated reaction area is in contact with an inhibitor, such inhibitor is unfeasible. For glucose reactive area, the rule is as follows

```
(2)     Activated_RA(?ra), Glucose_RA(?ra),
        Glucose_RA_Inhibitor(?i), contact(?i, ?a) ->
        unfeasible(?i).
```

The other states that if some non-activated reaction area is in contact with a reagent, the reagent is unfeasible[1]. For hemoglobin reactive area, the rule is as follows

```
(3)     Non-Activated_RA(?ra), Hemoglobin_RA(?ra),
        Hemoglobin_RA_Reagent(?r), contact(?r, ?a) ->
        unfeasible(?r).
```

Based on this model, an algorithm was developed to identify possible false-positive and false-negative results. It first creates an instance of urine containing instances of all substances of interest (i.e. all subclasses of reagents, confounders and inhibitors) and all reaction areas with their activation states. Then, the ontology is classified by a semantic reasoner. Given the rule (1), it is inferred that all instances of substance are contact with all instances of reaction areas. Moreover, given rules (2) and (3), instances of substances that are inhibitors for reactive areas whose instances said to be activated are classified as unfeasible – the same holding for instances of substances that are

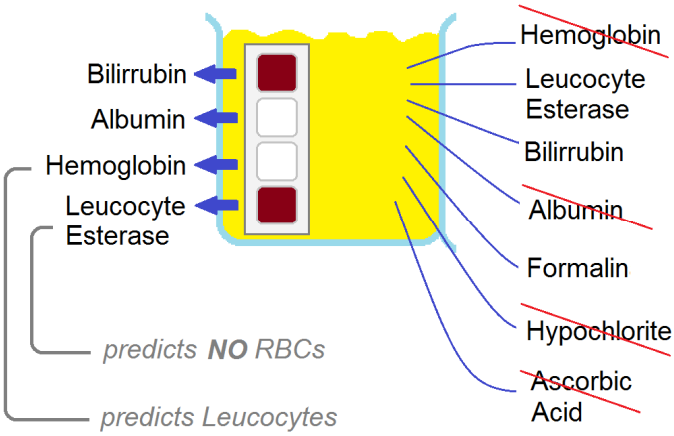
[1] Such rule is not completely correct since it is perfectly possible to have a reagent when a reaction area is inhibited. However, it is an approximation of reality that has practical value to make the inferences necessary during the task. Better representations are leaved as future work.

reagents for non-activated reactive areas. All instances of substances classified as unfeasible are discarded.

Following that, for each activated reaction area, the algorithm identifies its feasible reagents (i.e. instances of substances that are reagents for the reactive area and that were not classified as unfeasible). The result of the reaction area is accepted only if its analyte is among the feasible reagents. If, besides the analyte, there are feasible confounders, the algorithm raises suspicions about their presence, that are registered to be monitored during the microscopy phase. In order to allow future evaluation of suspicion, it is predicted what should be seen in microscopy until the end of the analysis to confirm the true-positive result.

Analogously, for each non-activated reaction area, the algorithm identifies the feasible inhibitors and, having any, raises and registers suspicions about them too, predicting which particles should be seen in microscopy if the result were positive – so that they must not be find in microscopy to confirm true-negative result. Particle prediction algorithms are discussed in sections 2.4 and 2.5.

Finally, masking identification was implemented in a simple and limited way. It simply compares colors of all reaction areas and, if all of them are similar to that of some coloring substance, masking by such coloring is suspected and ruled out the same way as the other false-positives.



ANALYTE	CONFOUNDER	INHIBITOR
BILIRRUBIN	-	-
ALBUMIN	-	-
HEMOGLOBIN	HYPOCHLORITE	ASCORBIC ACID FORMALIN
LEUCOCYTE ESTERASE	HYPOCHLORITE FORMALIN	ASCORBIC ACID

Figure II.6 - Dipstick analysis short example

A simplified example of the behavior introduced by the referred algorithm is presented in figure ii.6. There it is represented a urine sample containing a dipstick with only four reactive areas and some substances of interest (with the description of their roles regarding each analyte). Given rule (3), the presence of hemoglobin is unfeasible since hemoglobin reactive area is not activated – and so such instance is excluded from the urine sample. The same holds for albumin. Hypochlorite is a confounder for leucocyte esterase and hemoglobin reactive areas, so that both of them should be activated if such substance was present. As only leucocyte esterase reactive area is activated, hypochlorite is unfeasible and so discarded. Ascorbic acid is inhibitor for these same reactive areas and, as only that for hemoglobin is not activated, ascorbic acid is also unfeasible. The remaining substances are inferred to be feasible: bilirrubin and leucocyte esterase have their reactive areas activate as expected; and formalin is confounder for leucocyte esterase reactive area, which is activated, and inhibitor for hemoglobin reactive area, which is no activated.

Both activated reactive areas have their analyte as feasible reagents and their result can be accepted. Leucocyte esterase reactive area has a feasible confounder (i.e. formalin), what raises a suspicion about the correctness of such result. For this reason, leucocytes should be found in the sample to confirm the result. Formalin is also inhibitor for the non-activated hemoglobin reactive area, what raises a suspicion about this result too. It is predicted that RBCs should be found if hemoglobin is present in the urine. Thus, it must not be found in microscopy in order to confirm the negative result for hemoglobin.

This model of reaction areas environment is a simplification of its reality, since reaction areas can have several levels of activation. However, since there is no precise documentation about behavior of confounders over reaction areas, it was assumed that whenever they are present, they activate the susceptible reaction areas in any positive level – unless an inhibitor is present. Likewise, even though it is known that some substances just reduce the sensitivity of reaction areas, it is just considered that, when present, they simply keep reaction area non-activated, no matter how much reagent is in urine. Finally, since there is also no documentation about the exactly behavior of colorings on

reaction areas, they are assumed to, when present, coloring every reaction area with their colors.

A last note should be made on false-results (i.e. false-positives and false-negatives). Given the advance in dipstick technologies, the variety and likelihood of such situations are being reduced (e.g. ascorbic acid, that once was one of the main inhibitors and so deserve its own reaction area, nowadays is not such a problem given the increased quality of currently reaction areas, with its detector no longer present in dipsticks of most manufacturers). Nevertheless, although less frequent and coprehensive, it is still a conceptual possibility and a practical reality that may affect the quality of the analysis – deserving the effort made to deal with it.

2.3 Coherence Assessment

An incoherent urine sample was considered as any sample with two or more conflicting entities – e.g. particles, substances, pH, some patient information. To model such cases, it was added to the ontology two classes (presented in Figure II.7). “Conflicting Entity” is anything that conflicts with any other – which is, in turn, a conflicting entity too. “Incoherent Entity” is anything related to, at least, two conflicting entities conflicting to each other.

Provided the arguably infinite ways that something may conflict with some other, it is impracticable to define a general rule. Therefore, the incoherences were represented as SWRL rules stating the cases in which two instances related to some urine in a particular way conflict to each other. In these cases, the urine has each of the instances as conflicting entities. In Figure II.7 we have the example of alkaline crystal in acid urine. For this case there is a SWRL rule stating that if an urine has acid pH and contains some alkaline crystal, such crystal conflicts with the pH value and both are conflicting entities in urine. The rule is as follows

(4) `Alkaline_Cristal(?c), Urine(?u), Acid_pH (?pH), contains(?u, ?c), has_pH(?u, ?pH) -> conflictsWith(?c, ?pH), conflictsWith(?pH, ?c), hasConflicting(?u, ?c), hasConflicting(?u, ?pH)`

This way, having two conflicting entities (i.e. acid pH and alkaline crystal), such instance of urine sample can be classified as an incoherent entity.

The conflicts so far modeled also include presence of acid crystal in alkaline urine, presence of spermatozoon in urine referring to feminine patient, presence of *Trichomonas vaginalis* in urine referring to child patient (since this is a sexually transmitted parasite), glucose in urine referring to non-diabetic patient and non-crenated RBC in urine without hemoglobin.

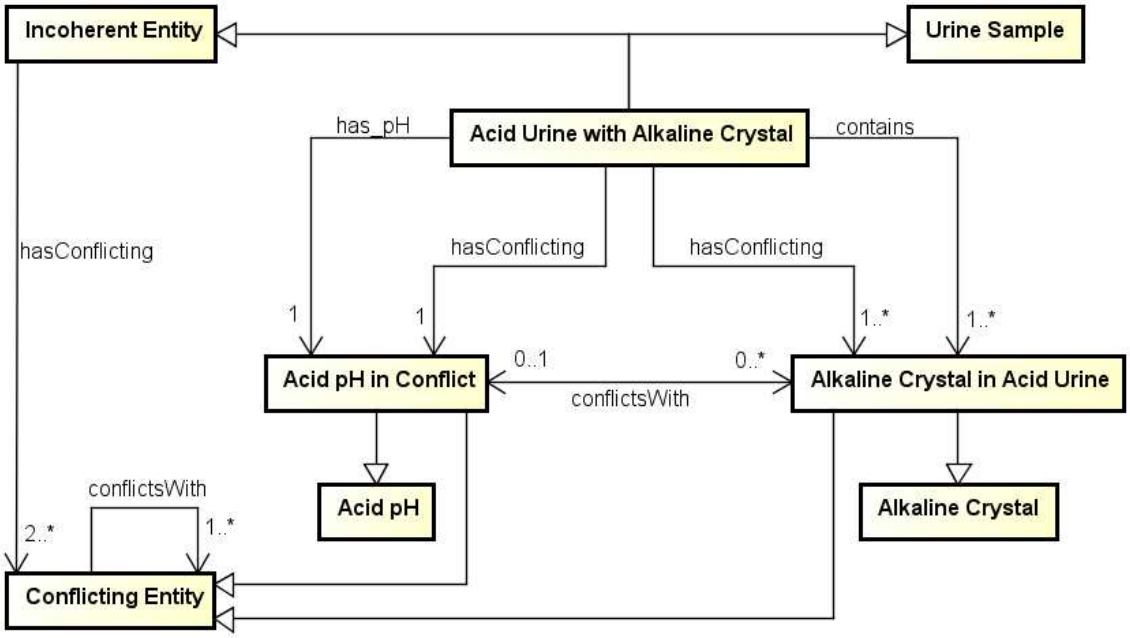


Figure II.7 - Ontological Model for Incoherences and Conflicts

Generally, urine samples are not really incoherent and this conclusion is a result of misinterpreting the findings – most of time confounding the type of an observed particle with a similar one. Then, in order to help in such cases, the possibly confusing types of particle were represented in the ontology using a "confounsinglySimilarTo" relationship, modelled using the Universal Relationship ontology design pattern [27]. Thus, if a class of particles may be confused with another class, any instance of the former will be linked to all instances of the later by a "confounsinglySimilarTo" relationship – and vice-versa.

Upon these representations, it was developed an algorithm to identify incoherent samples and what is conflicting in it, as well as suggesting what are

the confounding particles that may be being mistaken. The idea underlying the algorithm is that any particle that user reports as having been observed is either really observed or confused by some similar one. This way, the algorithm takes all conflicting entities (both contents and other features of the sample), and gather those that conflicts to each other, forming local conflicting groups. Then, for each group, it generates all combinations that keep one of its members and replace the others by confounders[2], when this is possible, creating local harmonic groups.

Following that, the algorithm generates all possible combinations of harmonic groups making up global harmonic groups (i.e. a set of local harmonic groups, each of them substituting its corresponding conflicting group). Then, it is created urine instances containing the entities (or being characterized by them, in case of features) of each global harmonic group, along with the remaining contents and features of the original sample that were not conflicting. Following, each of the urine instances are classified in order to verify if the new contents are conflicting. Finally, those instances that do not conflict are proposed as solutions for the incoherent findings reported by the user, along with the indication of tools suitable to confirm if such particles are in the sample indeed (the indication process is described in section 2.6).

For conflicts that do not involve particles (e.g. glycosuria in a sample of a non-diabetic patient), algorithm points out the conflict without indications of what might have been mistaken, just suggesting double checking dipstick and confirming the correctness of sample identification (in order to avoid sample referring to some wrong patient) or he clinical history and making contact with patient's physician (e.g. confirm the diabetic condition in case of finding glucose).

A short example of application of this algorithm is shown in Figure II.8. In this example, the sample have four conflicting things (1) that are gathered

[2] Currently, the only considered confusions are those related to particles. Among them: pyknocytes and spermatozoa; bacteria, amorphous phosphates and amorphous urates; air bubbles and isomorphic RBCs; yeasts in blastoconide phase and dysmorphic RBCs; hyalin casts and mucus. Confusions between other types of content or features of the sample will be researched in future efforts.

together in two local conflicting groups (2): urine referring to feminine patient conflicting with presence of spermatozoon and acid pH conflicting with the presence of amorphous phosphates (i.e. a type of alkalyne crystal). Then (3) the possible confounders for the members of the conflicting groups are identified: spematozoon may be mistaken with pyknocyte (i.e. a type of dysmorphic RBC) and amorphous phosphates may be confounded with amorphous urates (i.e. a type of acid crystal, that would not conflict with acid pH) or bacteria.

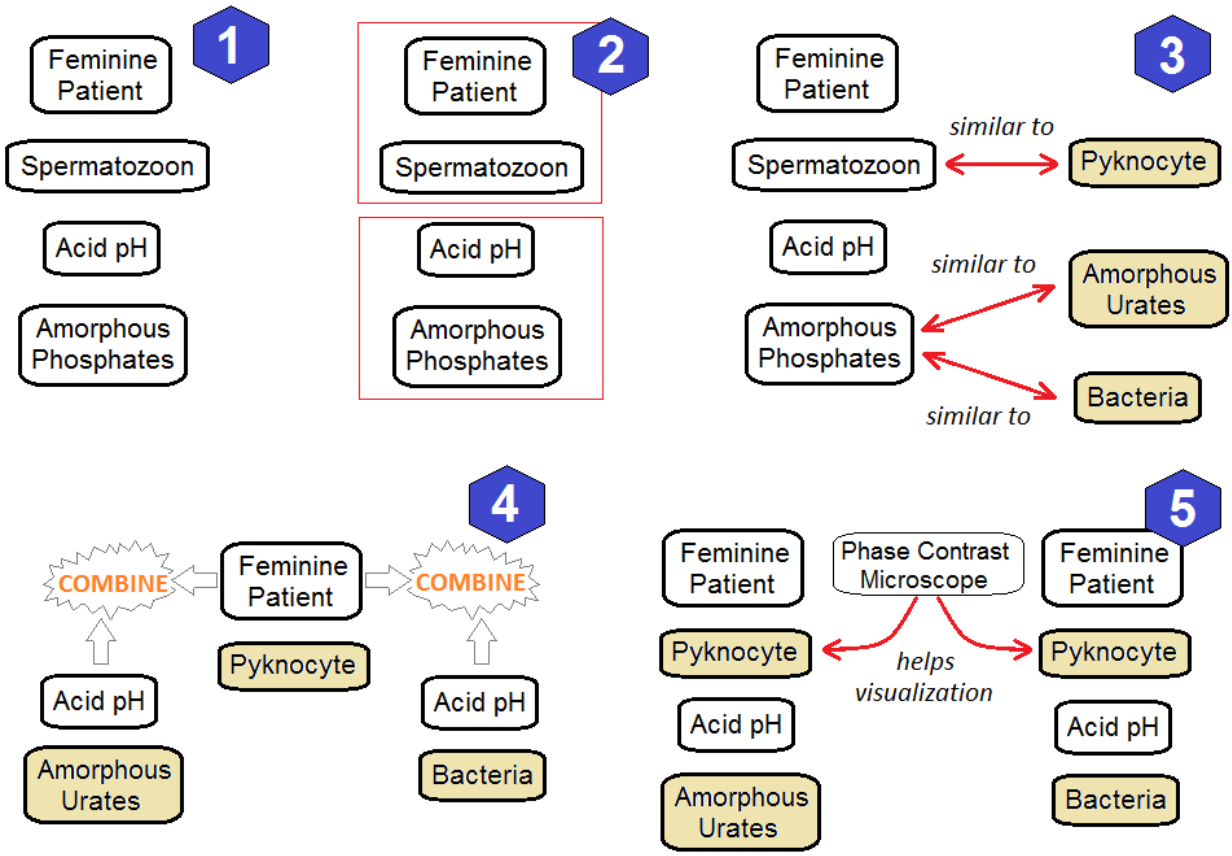


Figure II.8 - Coherence assessment short example

Following (4), one local harmonic group (i.e. feminine patient and pyknocyte) is created as an alternative for the conflict between feminine patient and presence of spermatozoon. Likewise, two harmonic alternatives (acid pH and amorphous urates or bacteria) are created for the conflic between urine with acid pH and presence of amorphous phosphates. The alternative for the first conflicting group is combined with those for the second conflicting group, generating two global harmonic groups (5). For this example, it is supposed that

no other feature or content of the sample conflicted with such global harmonic groups. So, as the contents and features of the created global harmonic groups do not conflict, the urine instances that contain them are not classified as an incoherent entity and both combinations can be proposed as solutions for the conflicts of the original sample. Finally, the algorithm identifies that phase contrast microscope may be useful to visualize pyknocyte and suggests it along with the solutions for the conflicts.

Besides the normal activity of the algorithm, if some special conditions are identified (e.g. too much glucose in urine occurs in cases of hyperglycemia and may indicate that patient is about to enter diabetic coma), system indicates conducts the user must take (e.g. if glycosuria is too high, immediately inform patient's physician). Currently, such indications are based on a set of rules directly coded in the system. Their ontological representation and unification with the module of selection of tools is leaved as future work.

In addition to the conflicts modeled as SWRL rules, coherence assessment may also identify other situations that should come to user's attention. One of them is the case of menstrual contamination, which is modeled by means of an urinary profile so that when the urine is recognized to present such profile, system raises the hypothesis of contamination and warns the user. Urinary profiles are further discussed in the next section.

A last coherence assessment the system carries out is the evaluation of suspicions raised during dipstick analysis. If it is not observed the findings expected to have been observed according to a positive result of a reaction area under suspicion, the system points out that the sample is likely to be contaminated and suggests new sample collection. Similarly, it gives the same suggestion if it is found something not expected due to negative result of a reaction area.

2.4 Profile-based Content Prediction

A urinary profile is a set of findings a urine sample should present – but not necessarily does. Therefore, it is not possible to say that a sample with a given profile has some specific contents, but rather that it is expected to have them. Considering that, an urinary profile is an abstract feature, existentially dependent of an urine sample, defined in terms of concrete entities (i.e.

particles and substances, whose existence is independent) whose presence is expected in the urine characterized by such profile.

Being a means of forecasting particles in the sample, it must not be necessary to identify all components of a profile to determine its existence – otherwise it would have no use since all that it could forecast would be what is already known to be in urine. Indeed, for each profile, there is a subset of findings that allows establishing its occurrence – some times even more than one (e.g. nephrotic profile can be determined either by strong albuminuria or by lipiduria).

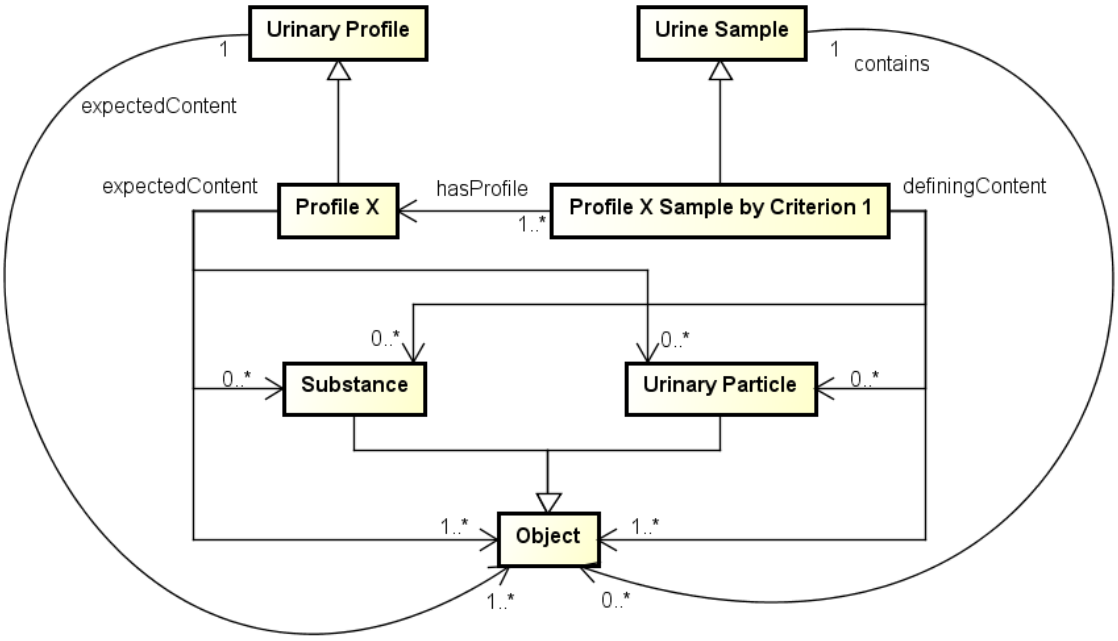


Figure II.9 - Ontological Model for Profiles

Based on this description of urinary profiles the ontological model presented in Figure II.9 was built. There, it is possible to observe that an urine sample may or may not contain some object (e.g. particle, substance) in it. There is also a class to represent the criteria[3] that a sample needs to meet to be characterized by some urinary profile. These criteria are defined in terms of the defining objects the sample must contain to feature the profile. By its turn, an urinary profile must refer to some objects that are expected to be find in a

[3] The use of criteria to classify samples was inspired by the "View Inheritance" ODP (http://ontologydesignpatterns.org/wiki/Submissions:View_Inheritance)

sample characterized by it. Moreover, it is from such objects that the defining contents of criteria classes will be taken. Finally, there will be possible several – at least one – criteria classes for each urinary profile, representing the different ways a sample may express it.

Figure II.10 shows the example of nephritic profile. There it is depicted the contents expected to be found in any urine characterized by such profile (related to it by the “expectedContent” relationship). Furthermore, it is represented two of the criteria by which a sample could be recognized as presenting nephritic profile – containing a RBC cast or strong hemoglobinuria and weak albuminuria.

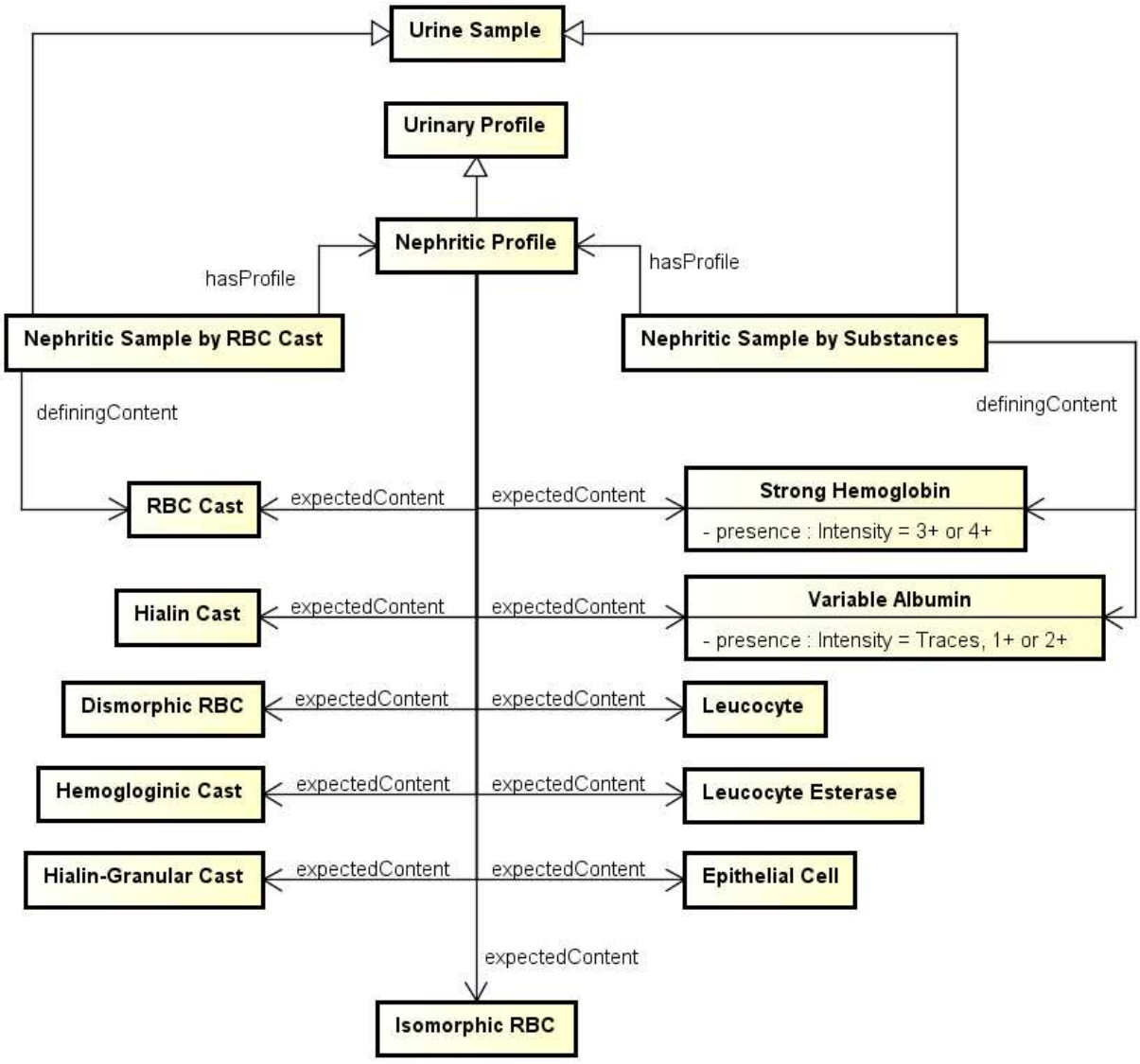


Figure II.10 - Nephritic Profile Example

Given such representation, we developed an algorithm to forecast sample contents based on urinary profiles. However, in order to take advantage from the capabilities of semantic reasoners for OWL 2, we represented this portion of ontology along with some additional particularities. First, considering that recognizing a urine sample as presenting some urinary profile will be used solely to forecast urine contents, it is not need to say that the sample has a specific instance of profile but rather that it presents any profile instance. Then, we created a prototypical instance (i.e. an ordinary OWL 2 instance but regarded as an ideal instance not intended to refer to particular known objects, but to some unspecific example of some concept) for every urinary profile. Following, we defined (by means of anonymous superclasses) that any urine sample instance that are classified as having a profile is linked to (i.e. *hasProfile*) the prototypical instance of such profile.

Likewise, it is not useful or even possible to say that a urine is expected to have some particular instance of particle (or other content), but rather that it will have any (yet unknown) instance of it. Thus, as for profiles, we created prototypical instances for every expected content given an urinary profile. The same way, we defined (also by means of anonymous superclasses) that any instance of a profile is linked (by *expectedContent* relationship) to each prototypical instance of the contents that make up the profile. Finally, by means of a property chain (i.e. `hasProfile o expectedContent -> expectedContent`) we defined that, if some urine sample has some profile and the profile has some expected contents, such sample has such expected contents (i.e. the instance of urine sample is directly linked – by *expectedContent* relationship – to those of the expected contents of the profile).

As a last note, we implemented similar mechanisms to represent the defining contents of a profile according to each criteria class. This way, the classes representing such criteria are defined as equivalent to the class of urines containing a combination of contents that allow he recognition of a urinary profile. Moreover, by means of anonymous superclasses, when a urine instance is classified as fulfilling a criteria class it is automatically linked (by *definingContent* relationship) to the prototypical instances of the contents that

allowed the sample to be classified under those criteria – and to be regarded as having the corresponding profile.

Provided such structure, we developed the following algorithm to predict urine contents based on urinary profiles:

1 – Classify the instance of urine
  1.1 – If it satisfies some criteria, it will be linked (via *hasProfile* relationship) to instances of the corresponding profiles and to the prototypical instances of contents that composes such criteria (via *definingContent*)
  1.2 – If linked to a profile instance, it will be linked (via *expectedContent*) to the prototypical instances of its expected contents
2 – Retrieve the prototypical instances of expected contents (i.e. linked to the urine instance through *expectedContent* relationship)
3 – Identify the classes of these instances
4 – Suggest these classes as things liable to be found in the sample
5 – For each criteria class the urine instance is classified within
  5.1 - Retrive the prototypical instances of defining contents (i.e. linked to the urine instance by *definingContent* relationship)
  5.2 – Identify the classes of these instances
  5.3 – Point theses classes as explanation on why the sample was classified as having the profile

Exemplifying the application of the algorithm, lets take an urine instance containing and RBC cast. As we see in Figure II.10, it would be classified as a “Nephritic Sample by RBC Cast” and so would be linked via *hasProfile* to the “prototypical nephritic profile” instance, via *definingContent* to the “prototypical RBC cast” instance and via *expectedContent* to similar instance of each of the

other contents indicated in the figure. Then, such instances would be retrieved, their classes identified and suggested to the user as content predictions – indicating that such forecast was possible due to the presence of a RBC cast in the sample.

The profiles currently represented within the system are described in Table 2. As exposed in the previous section, the menstrual contamination profile was not intended to help in forecasting findings, but to be used in incoherence assessment, indicating when hematuria should be regarded as menstrual contamination.

Table 2 - Represented urinary profiles

Profile	**Main expected contents and features**
nephritic	RBC casts, dismorphic RBCs, strong hemoglobinuria and variable albuminuria
nephrotic	strong albuminuria, lipiduria, epithelial cells and various kinds of casts
nephritic-nephrotic	about the sum of nephritic and nephrotic profiles
liver disorders	strong bilirrubinuria and urobilinogenuria, casts, RTECs
infection	microorganisms and leukocytes
renal infection	same as infection profile, but also presenting casts with microorganisms or with leukocytes
immunossupression	decoy cells or *Cryptococcus* yeasts, also presenting many other microorganisms
diabetes	glycosuria, ketonuria and yeasts
menstrual contamination	referring to a woman patient, with numerous squamous epithelial cells and isomorphic RBCs

2.5 Event-based Content Prediction

As it has been said, the knowledge of the events that may occur in urine can help to forecast which type of particles are likely to be found in a sample. As prescribed in UFO-B [28], events can be regarded as occurrents that take

place in a situation (i.e. its pre-state) and, after happening, bring about another one (i.e. its pos-state). Moreover, some continuants take part in the event, playing specific roles.

In urinalysis, the events of interest are those that take place in urine, both while it is *in vivo* or after collection, changing it in some way. So, a urinary event has as pre-state some situation in which a urine in some configuration is present and has as pos-state some situation with the same urine, but in a different configuration. Provided that, it was possible to recognize four types of urinary events: creation events, content change events, urine change events and destruction events – whose representation is depicted in Figure II.11. They subsume the concept of urinary event and have their participant objects indicated by relationships that reveal the role played by them.

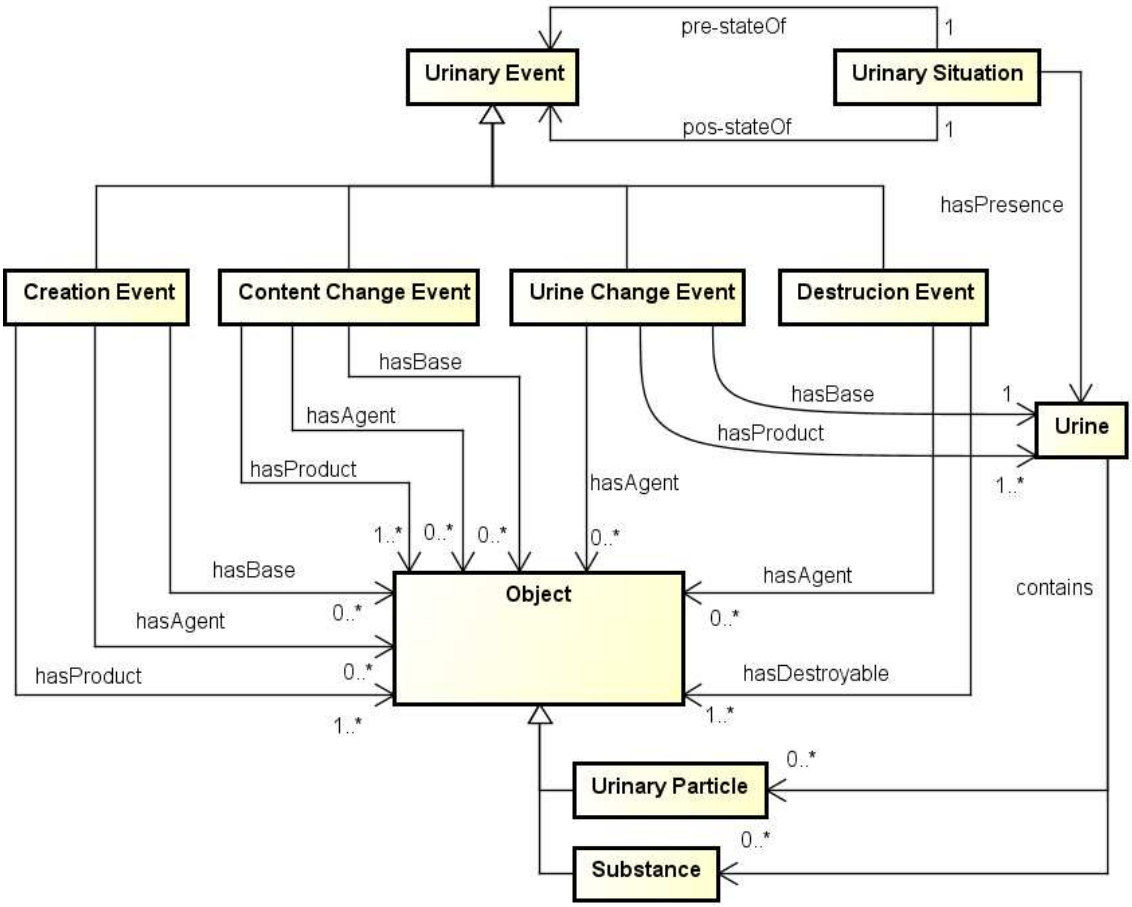


Figure II.11 - Ontological Model for Urinary Events

Creation events are those in which something that was not previously in urine appears. Thus, in its pre-state, urine does not contain the object and

began to contain it in its pos-state. Participant objects in a creation event may play one of three roles. The primary (and mandatory) is the product role – indicated by the “hasPoduct” relationship – and refers to what is being created by the event. Another one is the base role – indicated by the “hasBase” relationship –, which corresponds to the raw material the product was made from. Finally, there is the agent role – “hasAgent” relationship – that represents what operated the creation of the product from the base material. Every creation event must point out at least one product. However, to be useful, at least one of the remaining participants should be provided, in order to allow forecasting it in urine. Creation events may be just the appearance of a something (e.g. RTECs in urine *in vivo* due to bilirrubinuria) or transformation of some base material in something with a different nature (e.g. cast formation from Tamm-Horsfall protein).

Content change events are those in which something contained in urine changes but keeps its fundamental nature (e.g. RBCs go crenated). In its pre-state, urine contains the object in one form and has it in the other form in its pos-state. The roles played by participants are the same as in creation events – object in initial form plays the base role, changed object is the product and something that performs the change is the agent – and, again, product is the only mandatory participant.

Urine change events are similar to the content change events, except for the object that is in change is urine in itself (e.g. urine bacterial alkalinization). The similarity extends to the roles – urine plays the base role when in its initial form and the product role when changed, and the change instrument is the agent –, again with product as mandatory role.

Finally, destruction events are those in which something ceases to exist in urine, either by escaping it (e.g. ketones volatilizing at room temperature) or by being changed to something of different nature that is not identifiable or is of no interest (e.g. bilirrubin being destroyed by light exposure). These events may have only two participation roles: the destroyable (mandatory), played by the object disappearing during the event, and the agent, which is the object causing such destruction.

For each urinary event elicited during literature review and interview with expert, it was created a specific concept, subsuming its type of event. Along with that, concepts for situations corresponding to its pre and pos-states were also added to the ontology. Such situations were defined by necessary and sufficient conditions, so that an instance of a situation with some urine fulfilling these requirements could be automatically classified as pre/pos-state of some event. As there are events that only happen *in vivo*, pre/pos-states of these events have the presence of an *in vivo* urine as a necessary condition in order to classify a situation as such. Analogous requirements are made for states of *ex vivo* events.

Figure II.12 shows the example of nitrification event. As a creation event, it subsumes “Creation Event” concept. The situation of “Nitraturia and Bacteriuria” represents its pre-state. The urine in this situation has nitrates and a specific kind of bacteria (i.e. Gram-Negative). The situation of “Nitrituria” is its pos-state, which has the presence of some urine containing nitrites. The event happens with the bacteria (the agent of the event, as it is shown in the figure) degrading the nitrates (the base material) into nitrites (the product of the event). Although it is not represented in the model, the urine in the pos-state may also (and probably will) contain nitrates, bacteria and other components. Such components are not represented in order to keep the focus on the distinguishing features of the state (i.e. the presence of nitrites in urine). Besides, as this is an event that may occur both *in vivo* and *ex vivo*, there is no corresponding requirement about the origin of the urine that is present in pre/pos-states.

The same instance of urine may be simultaneously present in different situations, since the possible combinations of its findings may comply with their different sets of necessary and sufficient conditions. Thus, an instance of urine may be present in pos-states of some events and pre-state of others, linking them as in a graph. This structure mimics the chain of events that is assumed to exist in expert’s mind and so could be used to infer what may have hapened in urine and forecast its contents.

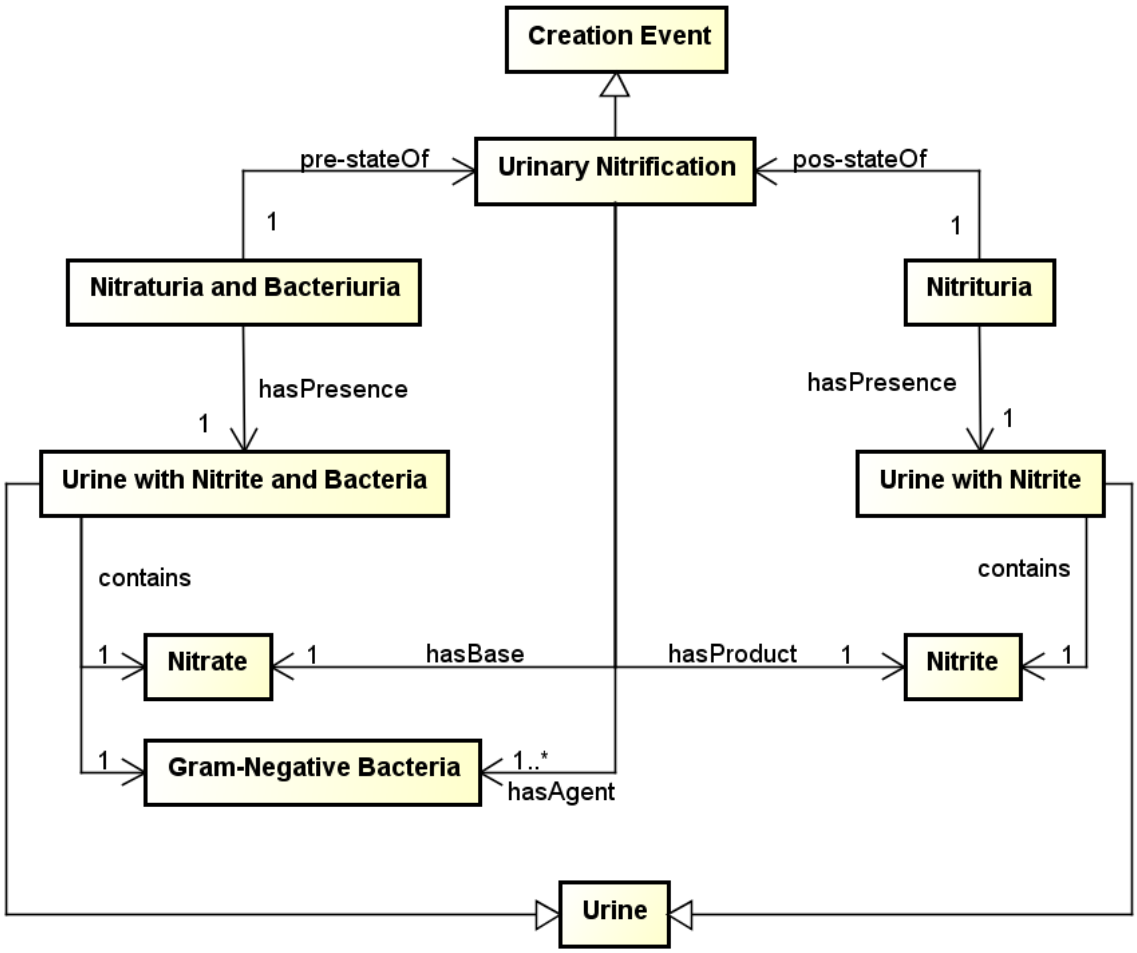


Figure II.12 - Example of Creation Event (Nitrification)

Similarly to the case of profile-based prediction, this portion of the ontology is intended to be used to investigate events that were not witnessed aiming to predict particles that were not yet observed. Then, with the purpose of allowing the reference to such hypothetical individuals, it was created a prototype instance of each event and each type of object that participates in the event. Each of the object instances was linked to that of the event they participate in through the relationship that denotes the role it plays. Classes of situations that represent pre/pos-states of an event were also linked directly to its prototypical instance (i.e. it was posed a property restriction indicating that all instances of such situations were linked to the prototypical instance of the event through “pre-stateOf” or “pos-stateOf” relationships, according to the case), so that any instance of these situations will be automatically linked to such prototype.

In view of browsing this event graph, we developed an algorithm upon this ontological structure. As a basic step, algorithm takes an instance of urine, creates a situation in which it is present, classifies the situation and identifies the events for which it is pre/pos-state. For events in which the situation is pos-state, the algorithm takes a regressive step, changing the instance of urine to meet the pre-state of the event (e.g. adding the agent that may have been present in the pre-state, causing the event). Likewise, for events in which the situation is pre-state, the algorithm takes a progressive step, changing the instance of urine to meet the pos-state of the event (e.g. adding the product that may have been created during the event). Every time an event is visited it is marked not to be processed again.

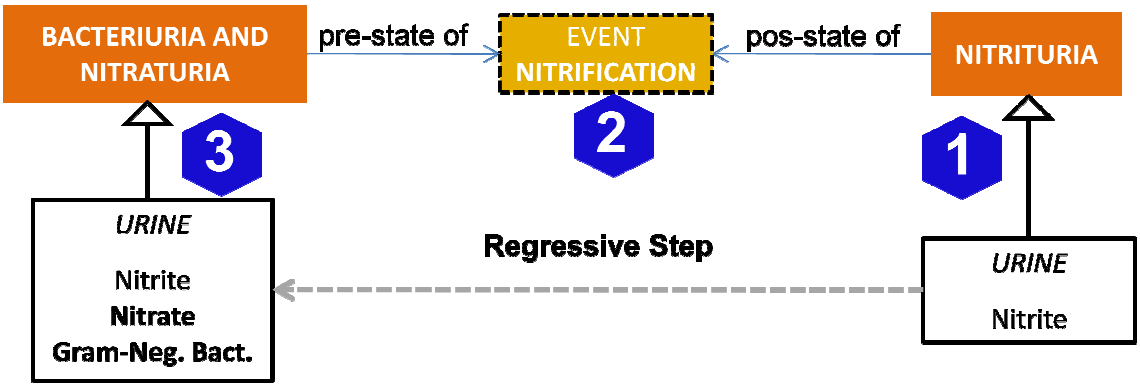


Figure II.13 - Example of regressive step

Figure II.13 presents a simplified example of a regressive step, based on the event presented in Figure II.12. Taking a situation in which a urine containing nitrite is present and classifying it, it will be inferred to be a situation of Nitrituria (1). Then, being a Nitrituria, it will be classified as pos-state of Nitrification (2), which has Nitrate as its base and Gram-Negative Bacteria as its agent. Thus, the instances of these to urine contents are added to the contents of the urine instance (3), changing the urine in a way that it would be classified as the pre-state of the event (i.e. Bacteriuria and Nitraturia).

Based in these basic steps, a longer algorithm was devised to browse the event graph proposed, as following:

1 – Create an instance of situation in which the instance of urine under analysis is present
2 – Classify the situation
  2.1 – If classified as pos-state of a non-visited event
    2.1.1 - Take a regressive step
    2.1.2 – Classify it again and return to 2.1
3 - Create a situation with an *in vivo* urine instance with same contents and features of the sample under analysis
4 – Repeat step 2 (and its substeps), but with the new situation; then, follow to step 5
5 – Classify the situation
  5.1 – If classified as pre-state of a non-visited event
    5.1.1 - Take a progressive step
    5.1.2 – Classify it again and return to 5.1
6 – Create a situation with an *ex vivo* urine instance with same contents and features of the *in vivo* specimen
7 – Repeat step 5 (and its substeps), but with the new situation, then, follow to step 8
8 – Repeat steps 2 to 7 until the situation is not classified as some non-visited states during such steps

While the algorithm is processed, the succession of events that were visited (i.e. the succession of events that may have happened to the urine), as well as what was changed in urine, is recorded to be used to explain system reasoning to the user. It includes the objects that participated in the event (e.g. agents, base materials), what will be used as source for prediction of particles.

The idea underlying this algorithm is consider what may have happened to the urine since its creation until the moment of its analysis. Having a sample, imagine all possible events that may have happened to the sample since it was collected, (i.e. step 2). Then, after modifying the sample instance to have a condition similar to that right after its collection, assume it is still inside the patient (i.e. step 3) and imagine all possible events that may have happened to the sample since its production (i.e. step 4). Then, modifying the urine to reflect that possible initial state, imagine all the events that would happen inside the

patient if the urine was in such condition until a moment before the specimen collection (i.e. step 5). Following, modifying the sample accordingly, assume that it was collected with such possible conditions (i.e. step 6) and imagine all the events that could have happened until the analysis of the sample (i.e. step 7). Repeating such cycle again and again until no more event is left to consider (i.e. step 8) would increasingly provide further insights about the changes in the urine and thus about its possible current contents.

Table 3- Currently represented urinary events (**a**=agent; **b**=base; **d**=destroyable; **p**=product)

Pre-State	Event	Pos-State
Tamm-Horsfall protein (***b***) in *in vivo* urine	Cast Formation(***creation***)	Casts (***p***) in *in vivo* urine
Microorganisms (***a***) in *in vivo* urine	Immune Response (***creation***)	Leukocytes (***p***) in *in vivo* urine
Bilirrubin (***a***) in *in vivo* urine	Secretion of Tamm-Horsfall protein (***creation***)	Tamm-Horsfall (***p***) in *in vivo* urine
Bilirrubin under light (***d***) in ex *vivo* urine	Bilirrubin Destruction (***destruction***)	Ex vivo urine with no bilirrubin
Urobilinogen (***d***) in ex *vivo* urine under light (***a***)	Urobilinogen Destruction (***destruction***)	Ex vivo urine with no urobilinogen
Glucose (***d***) in ex *vivo* urine at room temperature (***a***)	Glucose Destruction (***destruction***)	Ex vivo urine with no glucose
Ketones (***d***) in ex *vivo* urine at room temperature (***a***)	Ketones Volatilization (***destruction***)	Ex vivo urine with no ketones
RBC (***d***) in urine	RBC Lysis (***content change / creation***)	Lysed RBC (***p***) and hemoglobin (***p***) in urine
RBC (***b***) in dense urine (***a***)	RBC Crenation (***content change***)	Crenated RBC (***p***) in urine
Leukocyte (***b***) in urine	Leukocyte Lysis (***creation***)	Leukocyte esterase (***p***) in urine
Lipid (***b***) in urine	Cholesterol Crystal Formation (***creation***)	Cholesterol crystal (***p***) in urine
Urease producing bacteria (***a***) in acid/neutral urine (***b***)	Bacterial Alkalinization (***urine change***)	Alkaline urine (***p***)

Table 3 presents the events currently represented in the ontology. The types of events and roles of the urine contents (***a*** for “agent”, ***b*** for “base”, ***d*** for “destroyable” and ***p*** for “product”) are between round brackets.

2.6 Selection of Tools

Besides dipstick, the tools that professionals may use in urinalysis are basically those employed during microscopy (i.e. microscopes, specific reagents and urinary colorings) either to make it easier to recognize a particle or to differentiate it from another confusing particle type. Such tools do so in three main ways: (i) providing better visualization of particle details, (ii) revealing particle aspects otherwise imperceptible or (iii) interacting with a particle so that it behaves in a particular way.

In (i) the details that may be highlighted are many and vary from particle to particle. Moreover, it is hard to define how a tool can emphasize them. Therefore, it is only possible to say that tools that work as in (i) facilitates the visualization of particles in an unspecific way.

On the other hand, tools that acts as in (ii) allow the specification of which aspects are revealed. Then, it is possible to say that when some particle is observed using such tools, it can be further described by a new aspect previously not known. Moreover, as revealing a new aspect of a particle can also be considered providing a better visualization of such specific detail, it is possible to say that any tool that acts as in (ii) also facilitates the visualization of some particle.

Finally, acting as in (iii) involves causing a behavior that is not peculiar to the particle. It may be manifested as in (ii), causing a temporary visual or behavioral aspect (e.g. brighting, movement), what is liable to be represented the same way. In other cases, it may happen structurally changing particle (e.g. destroying it). These interventions would be richer and better represented through events. However, as the task requires nothing more than a tip on what to use to identify or differentiate a particle, there is no need of descriptions of participations roles or pre/pos-states. Thus, just representing the relationship between the tool and the particle is sufficient for the objectives of this work.

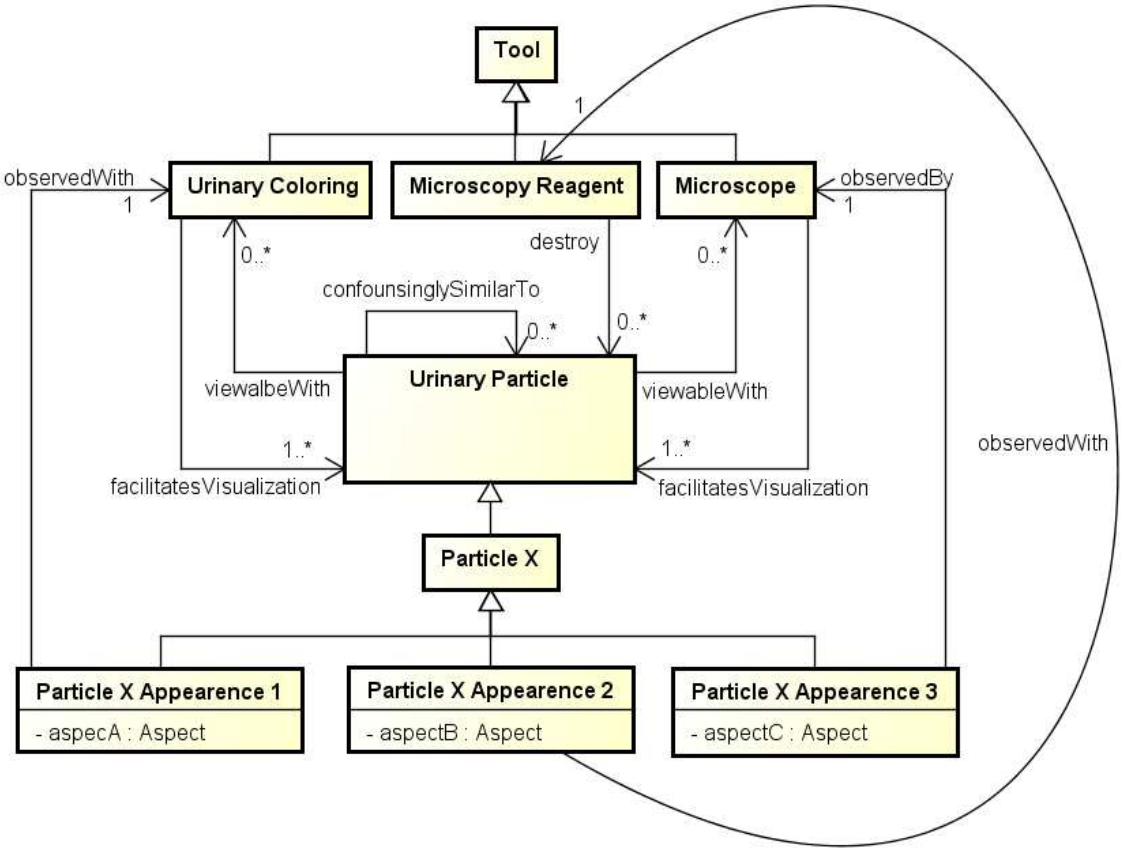


Figure II.14 - Tools Ontological Model

With such considerations in mind, it was constructed the model presented in Figure II.14. As it is modeled there, any microscope facilitates the visualization of some particle. However, some particles present different aspects when observed by some types of microscope (e.g. lipid droplets bright when seen trhough polarized light microscopes), what is represented by the “obsevedBy” relationship. Urinary colorings causes similar effects (e.g. nankin in the sample reveals the capsules that wraps *Cryptococcus* yeasts), what is represented by the “observedWith” relationship, and doing so can also be considered to facilitate visualization of these particles. Specific reagents can act similarly (e.g. acetic acid enhances nuclear detail of some leukocytes), what is represented by “obseredWith” relationship, or as in (iii), what is exemplified by the relationship “destroy” (e.g. acid acetic destroys amorphous phosphates and RBCs). As some tools may allow visualization of particles that otherwise would not be possible to see (e.g. lysed RBC are only observable through phase contrast microscope), such condition is represented by the relationship “viewableWith”, indicating for each particle type what kinds of tools allow

visualizing them. Finally, as there are types of particles that may be confused with others, it is represented by the relationship “confounsinglySimilarTo”, that may relate a type of particle to other.

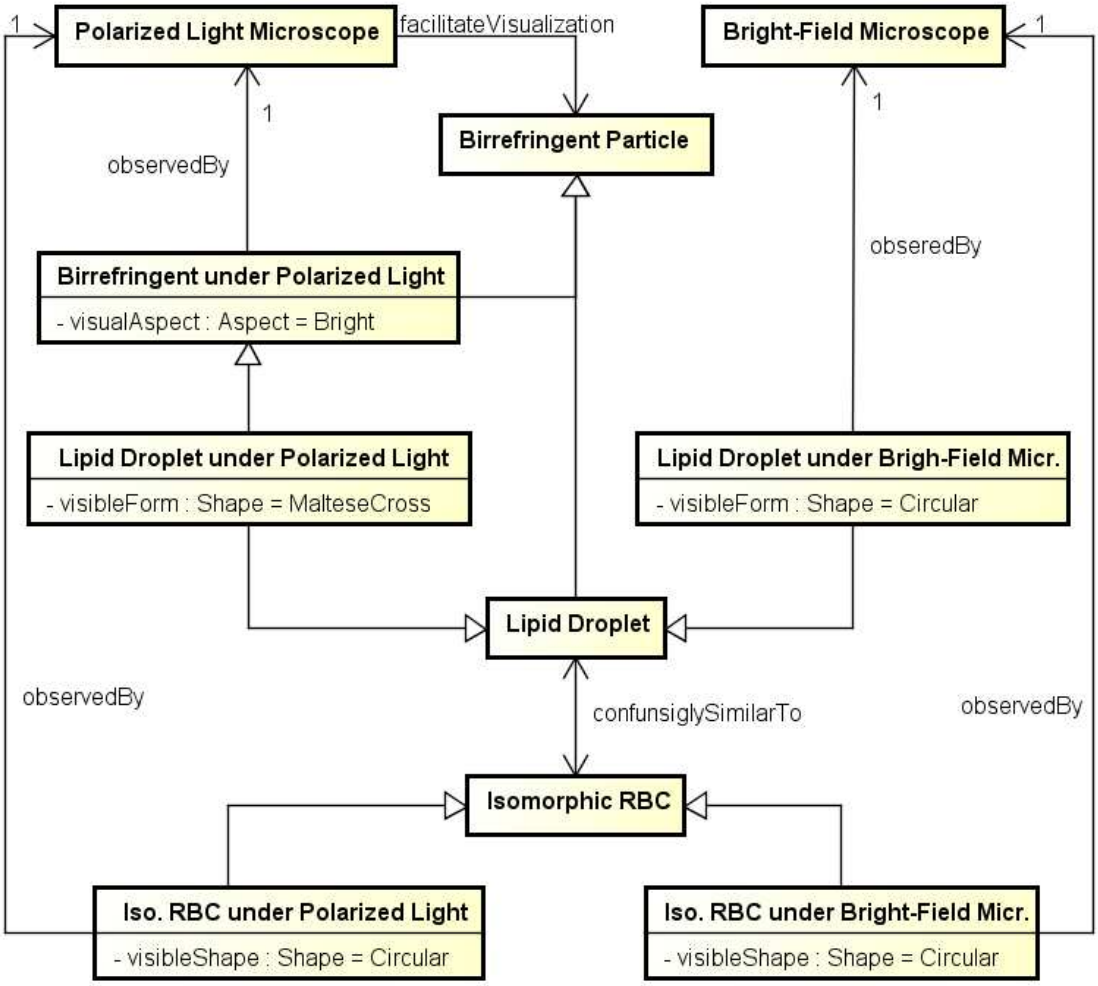


Figure II.15 - Example of Tool (Microscopes)

Figure **II.15** brings the examples of bright-field microscope and that with polarized light. As it can be seen, only polarized light microscope facilitates the visualization of birefringent particles, doing so by making it shine (i.e. birefringent under polarized light feature a bright aspect). Considering lipid droplets and isomorphic RBCs, if both are observed trough bright-field microscope, both will appear as circular particles, being confusingly similar to each other. But, if observed under polarized light, lipid droplets will shine (given that they are birefringent particles) and appear with Maltese cross shape, while isomorphic RBCs will remain as circular particles. Given this difference in

behavior of these particles, polarized light microscope could be indicated to differentiate them

In view of using this representation to provide advise to the user on what to employ to observe a particle, a simple algorithm was developed. If the aim is just observe the particle, the algorithm simply take the instance of particle and classifies it. According to the type of the particle, the algorithm retrieves the classes of tools the particle is “viewableWith” and suggests to the user. Moreover, it creates an instance of each tool that allows visualization of the particle and establish an *observedWith* (or *observedBy* for microscopes) relationship between it and the particle instance. Then, if the particle is classified in a class of particular appearance – thus being described with a new feature – such feature is reported to the user along with the tool, indicating the behavior of the particle when using the tool.

In case of special actions of microscopy reagents, when an instance of the reagent is created and the particle is classified it is automated linked to the reagent by the relationship that represents the interaction between particle and reagent (since such relationships are represented using the Universal Relationship ODP). Then, the interaction represented by the relationship is reported along with the reagent, also indicating what will happen to the particle when using it.

Finally, if the aim is to differentiate confounding particles, it may be provided two particle instances, one of each type. Following that, the previously described process is carried out, but only the tools that causes different behaviors for each particle are reported. Alternatively, it may be provided only one particle. In this case, it is created an instance of each particle type the particle may be confused with. Then, each of then is paired with the provided particle and the before mentioned process is carried out.

Currently, the ontology models only six types of tool. Polarized light microscope makes birefringent particles (e.g. lipid droplets, some crystals) show a bright aspect. Phase contrast microscope allows the visualization of some particles (e.g. lysed RBCs). Nankin allows the identification of the capsules that wraps *Cryptococcus* yeasts. Papanicolau enhance details of decoy cells.

Finally, acetic acid destroys some particles (e.g. RBCs, some alkaline crystals) and enhance details of others (e.g. leukocyte nucleus).

2.7 System Dynamics

After constructing each of the described ontological models and algorithms, all was tied up in a prototype of knowledge-based system whose complete dynamics is presented in Figure II.16. Everything begins with the user (i.e. someone performing the analysis of a sample guided by the system) providing patient information (i.e. age, gender, clinical conditions) and the result of a dipstick test. This information is submitted through an interface indicating, for each reaction area, the possible values it can present and the possible information that can be provided about the patient. Then, the system performs the dipstick analisys. For every possible false-positive or false-negative results, system registers the corresponding suspicion to be monitored during coherence assessment task. If there is some unacceptable result, system informs the user and the guide session ends. If all dipstick results are acceptable, system checks the coherence of the findings (at this point, solely between dipstick results and patient information) and warns user about possible incoherences, suggesting conducts. The conflicts reported at this point may be corrected by the user, but do not cause the abortion of the guidance session if remain unaltered.

Following that, user informs the particles found on a microscopic field and system assesses their coherence (among themselves and in relation to dipstick results and patient's information). If there are incoherent findings (e.g. acid crystal in alkaline urine, spermatozoon in woman's sample), system informs the user and points out which particles may have been confused,

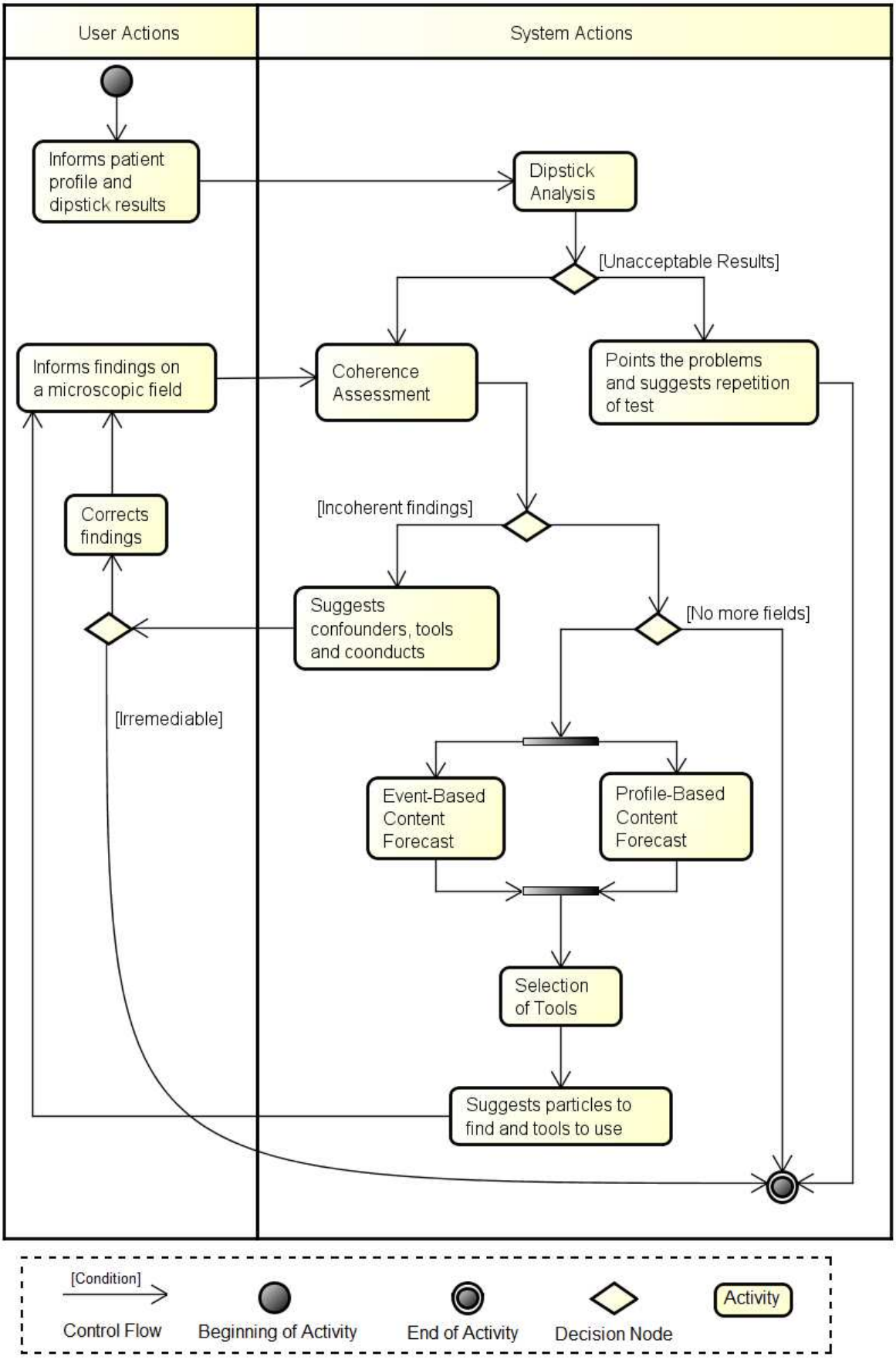


Figure II.16 - System Dynamics

indicating tools to better identify them selection of tools). If the incoherences involves dipstick results that are under suspicion, system indicates that possibility. At this point, user can change the findings reported (either from dipstick or microscopy). After correcting findings, system assesses their coherence again. If the conflicts cannot be corrected (e.g. user keep affirming that those are the findings), it is possibly the case of an odd sample that is beyond the knowledge of the system. In such cases, the guidance session ends, with the suggestion of repeating the analysis with a new sample.

If findings are coherent, system forecasts particles based on profiles and events that may have been happened in urine. At this point, system explains why such particles were predicted, both by informing the profiles the sample presents (along with the criteria use to identify such profiles) and the chain of events that may have happened to the urine. Based on what was foreseen, it selects tools to help user on the analysis of the next microscopic field. Then the user observes one more field, reports what is observed and the cycle repeats. The guidance session ends when there are no more microscopic fields to be observed. If some contamination suspicion or unsolvable conflict remains at the end of the session, the system indicates so and suggest repetition of the exam.

3 Results and discussion

The aim of the system developed in this work is modelling urinalysis knowledge in order to give advice on what to do while conducting a urine examination. Provided that, it should be evaluated measuring in which extent it behaves as an accredited professional would. Yet, considering expert's superior performance and that great part of the knowledge of the system is based on his own, it seems to be the right decision to base the evaluation on him. Such choice is possible since he did not directly take part on the development of the system, with his influence limited to the knowledge elicititation during the series of interviews and the elaboration of the set of examples used to evaluate the system.

The expert elaborated 17 fictitious examples of urine sample, through the filling of an electronic form with three sections. First section includes patient's information (i.e. gender, age and specific clinical conditions) and

dipstick results and ends asking what the information reported so far would make the expert expect to find in microscopy.

The next section asks for microscopic findings. It includes the amount of found RBCs (gathered by type as isomorphic, dismorphic, crenate, turgid or lysed), epithelial cells (gathered by type as squamous, transitional and RTEC) and leukocytes and the occurrence of casts (by type), crystals (by type), bacteria, yeasts, other microorganisms (e.g. *Trichomonas vaginalis*), artifacts and any other particles. The section ends asking what tools (and what for) expert would have used during the examination of the particles, which particles could be confused with some other type (and which one).

The last section asked which combination of findings (both from dipstick and microscopy) would have led expert to search for some specific particle, if the provided example of sample represented any special situation, which conducts the expert would have adopted provided the findings on the exam and which combination of findings would have triggered such conducts

Each example sample was submitted to the system during a guidance session. To enable that, the findings in microscopy of each sample were randomized, for each guidance session, so that the reported amounts (or occurrences) of particles appear distributed along 10 microscopic fields. Then, for each sample it was carried out three complete guidance sessions, that began by providing patient's information and dipstick results and then the findings for each of the 10 microscopic fields. The advice given by the system during the session was registered to be evaluated against (i) the effective findings in the example, in order to evaluate if the system managed to forecast them, and (ii) the expert's answers at the end of each form section, aiming to evaluate if the system suggested tools and conducts suitable for the case as well as correctly related the findings in the urine. The system was evaluated in four axis, one for each of the main modelled tasks – except for profile-based particle prediction and event-based particle prediction, that were gathered under a unique particle prediction category.

3.1 Results of Dipstick Analysis

Since none of the samples covered cases of contamination of the sample by substances that might have influenced dipstick results, the dipstick

analysis was not quite demanded. This prevented the complete assessment of this task, which represents a limitation of our evaluation method. However, it was still possible to observe its detector dynamics, raising expected contamination suspicions that were later correctly dismissed.

3.2 Results of Coherence Assessment

Coherence assessment was another task not completely tested by our evaluation approach. Since the examples were elaborated as if they were the underlying reality of the sample, instead of the perception of an inexperienced observer, almost no conflicts were introduced with the described findings – except for the presence of spermatozoon in a sample of a woman (which was identified by the specific SWRL rule previously presented) and a possible case of menstruation contamination (recognized by matching the requirements for menstrual contamination profile), both of them causing the suggestion of repeating the exam on a new sample.

Even so, it was possible to partially asses its correctness by not pointing incoherences in such harmonic samples (e.g. all crystals were in correct pH conditions, not causing any warnings) as well as by correctly ruling out the suspicions about sample contamination. Regarding the suggestion of exernal conducts, those about new sample collection were correctly handled. However, those requiring send for patient's physician were not triggered due to be caused by too specific reasons, overlooked during system development – what is not a big deal, since such rules can easily be added with no side effects on the rest of the system, provided that those are just consultative conducts, not causing the abortion of the analysis. Being probably the most ad-hoc decision in the analysis (i.e. that does not have a wide support in literature and may vary even from laboratory to laboratory), a higher failure level would be expected in this task – what does not relieve it from future study to improve system suggestions.

3.3 Results of Particle Prediction

Regarding the way that the requirements in [17] were posed, particle prediction seems to be the most important task on the analysis – and so will receive further attention. It was assessed in two distinct moments: right after

dipstick test (when it is only known its results and the basic patient's profile) and during microscopy. Evaluation was done by calculating precision and recall for its predictions (since the system just give positive forecast, specificity, negative predictive value and accuracy are not applicable). To make such calculations, it was considered the number of effective occurrences of particle types *versus* the predicted occurrences. By this measure, the presence in the sample of both 100 leukocytes and just 1 would count as the occurrence of only one particle type (i.e. leukocyte). Expert's particle predictions were considered the gold standard values when forecasting particles right after physicochemical test. Effective occurrences of particle types – as established in the sample description – were the gold standard in predictions during microscopy.

On predicting particles based solely on dipstick and patient's profile information, expert predicted 54 occurrences along the 17 example samples, while the system predicted 91, with 49 results coinciding with expert's predictions. This represents a precision of 53.85% (i.e. 49/91) and a recall of 90.74% (i.e. 49/54). On average, precision was 59,78%, with σ = 22.17%, and recall was 91.42%, with σ = 18.66%. During microscopy, the 17 samples presented altogether 143 particle type occurrences and the system predicted 182, with 113 hits. This results in a precision of 62.08% (i.e. 113/182) and a recall of 79.02%. On average, precision was 61.09% per sample, with σ = 14.48%, and recall was 75.43%, with σ = 14.35%.

Relatively low precision in particle prediction based on dipstick and patient profile is probably due to moderate expert prediction. Such hypothesis is raised on the grounds that, if considering as gold standard to this phase the actually findings in microscopy, the precision would rise to 68.13% (i.e. from the 91 predicted particles, 62 were found on microscopy). It might be a sign of further internalization of knowledge packaging by the expert (e.g. for him it is so obvious to expect some findings that it seems not necessary to mention).

Overall relatively low precision rates are also due to other factors, such as forecasting of all particles of a profile when a sample is classified as presenting such profile and the excessive premissivity in predicting casts and some other particles (e.g. existing a cast and a particle, system invariably foresees casts containing such particle). This suggests the need of fine tuning

these predictions (e.g. defining further restrictions to recognize pre-state of specific casts formation, identify subdivisions of urinary profiles).

Recall rates during microscopy are lowered by impossibility of predicting crystals. Since most of them have no clinical significance it does not seem to be possible to correlate them with other findings, so that in most of cases, the only treatment that is possible is their differentiation and coherence assessment to avoid reporting wrong crystals. Beyond that, hyalin casts, as well as little amounts of leukocytes and RBCs, may appear in a variety of normal conditions, being also unpredictable in such cases. One possible solution for that would be establishing some sort of normal urine profile, that could be used to allow foreseeing such particles that could be found in urine of normal patients.

It is noteworthy that, although the flaws in predicting some findings, the proposed models seem to have demonstrated their strength. Indeed, all the findings that were not predicted would be so with the addition of some pieces of information (e.g. events, profiles). Although not yet present in the prototype, such additional information is perfectly suitable to the knowledge models already defined, showing their success in representing the domain – probably the most important contribution of this work.

Moreover, the results are pretty good when compared to the brute force alternative (i.e. when no pondered prediction is done, considering that any particle may be present in any sample). In each example sample the system predicted about 11 particles during microscopy phase. If it had naively predicted the 54 real urine components represented in ontology (i.e. all cells, all crystals, all casts, and so on) it would obviously reach 100% recall, however, at the cost of poor 15.28% of precision.

3.4 Results of Selection of Tools

Finally, regarding selection of tools, the expert suggested the use of specific tools 40 times along the 17 samples – always recommending one of the three types of microscope represented in the ontology (i.e. bright field, phase contrast and polarized light). In this aspect, the system suggested tools 44 times – exactly the same 40 expert's suggestions plus 4 recommendations of nankin and Papanicolau stains when expecting the appearance of

Cryptococcus and decoy cells, respectively. Notably, the system justifications for using the tools also coincide with those of expert, plus a dozen of others related to excessive prediction of some particles.

As a last note on evaluation, the amount of examples used to assess the success of the prototype was pretty small and it should be further tested against a greater set of cases with broader coverage of the situations urinalysis professionals may come across. Such new testing round would help in filling the gaps left by this first evaluation, better proofing the capabilities of the prototype in all activities as well as answering the questions and hypotheses raised during the analysis of the present results.

4 Conclusions

The objective of this work was developing a prototype of a knowledge-based system for urinalysis in order to demonstrate how the knowledge of this domain can be computationally represented and processed in order to allow a system to reason over urine samples as a professional would. For this purpose, knowledge about urinalysis was elicited by reviewing literature and conducting a series of interviews with an expert on the field, being identified five main activities performed during the analysis of a sample. From such knowledge, it was built a general ontology for urinalysis, complemented by special ontological constructions to deal with specificities of each of the identified activities. The system itself comprises this ontological knowledge and algorithms built upon it and use them to reason over information about a urine sample. It was evaluated against expert's knowledge by confronting it with examples of urine samples that, in his opinion, represent important cases that a urinalysis professional should be able to deal with.

The prototype showed interesting ability to deal with the cases submitted to it during evaluation. It was able to forecast most particle occurrences said to be reasonably expectable given initial data (i.e. dipstick results and patient information). Similarly, it had great success in foreseeing the particles that effective made up the urinary sediment of the examples by considering the findings as they were being reported – doing such reasoning consistently with the heuristics suggested by the expert for each example.

Every tool the expert stated as useful for the cases was also recommended by the system.

The main limitation of the system is the prediction of the number of particles that should be expected in the sample (i.e predictions are made only in terms of particle occurrences). Besides some profiles that are partially defined in terms of the amount of certain types of particles present in the sample and the direct relation between some substance levels and the number of corresponding particles normally present in urine (e.g. hemoglobin and RBCs), there is not much record about associations between different quantities of findings. Even in expert experience it is not evident the existence of clear ever-present patterns concerning this aspect.

Other noteworthy issue is the fact that the system tends to forecast more findings than it is normally find. It is still a little slice of all possible findings that may be present in a sample, what means a great heuristic filter. In spite of that, this may be sign of a yet not optimized process of hypothesis formulation. Anyway, though it must be carefully studied to further improvement, it is not so big deal since, as stated the expert in one of the interviews, the important is bearing in mind all the likely scenarios in order to avoid missing important findings by not considering some of them.

This work opens some opportunities for future work. The most immediate one is turning the developed prototype into a fully functional system. Since the work focused on the knowledge representation and reasoning, the prototype was delivered as a limited old-fashioned desktop application. This way, the design of appropriate interfaces for the system to enhance interaction with the user as well as its adaptation to other plataforms (e.g. use through web, mobile applications) are important next steps. After such improvements, it should be carried out a further level of evaluation, comparing the performance in anlyzing real urinary samples between people using the system and others not using it.

Beyond the purpose of guidance, new interfaces may be coupled to the system developed in this work with the aim of allowing new uses of its intelligent core. Among such uses, it may be developed a simulator of urinalysis. In such simulator, a sample would be represented by its dipstick results and a sequence

of pictures of microscopic fields, as well as an underlying description of it, using ontology concepts to describe the findings of each picture. With that, just like the current prototype, the simulator would guide the user during the analysis of the fictitious sample. Other interfaces may also allow directly querying the knowledge represented in the ontology, as well as testing user's hypotheses (e.g. is tool X suitable to differentiate particles A and B?).

Besides that, new rounds of meetings with the expert may be interesting to enlarge the coverage of the system. Such enlargement may happen by introducing additional knowledge (e.g. events, profiles, tools) in order to solve what was missed by the system during evaluation or covering cases not yet modeled. Along with that, it may help to fine tuning the knowledge already represented, aiming to lower the number of excessive predictions that occurred in some samples. Additionaly, research may be conduct towards the issue of better predicting the quantities of findings in urine.

Finally, considering the existence of several other clinical laboratory tests (e.g. blood, cerebrospinal fluid), the developed prototype may give insights about how to model such domains. For example, there are certainly events that happen in their samples that influence their contents and it is probable that they may be characterized by structures that resemble urinary profiles (i.e. sets of findings that tend to appear together). Likewise, it is reasonable to expect that some kind of detector is employed in those analyses – that would probably be analogous to urinary reaction areas – as well as their particles may be observed by microscopes, whose function and purpose must remain pretty much the same from domain to domain.

5 List of abbreviations

API – Application Programming Interface

KB – Knowledge-Based

KBS – Knowledge-Based System

ODP – Ontology Design Pattern

OWL – Web Ontology Language

RBC – Red Blood Cell

RTEC – Renal Tubular Epithelial Cell

SWRL – Semantic Web Rule Language

6 References

1 - Strasinger SK, Di Lorenzo MS. Urinalysis and Body Fluids. 5th ed. Philadelphia, USA: F. A. Davis Company; 2008.

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

3 - Fogazzi G B. The Urinary Sediment: An Integrated View. 3 ed. Milan, Italy: Elsevier; 2010.

4 - Gruber TR. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition. 1993;5;2:199-220.

5 - Guarino N. Formal Ontology and Information Systems. In: Proceedings of International Conference on Formal Ontology in Information Systems (FOIS'98). Amsterdam, Netherlands: IOS Press; 1998. p. 3-15.

6 - Yang Y, Calmet J. OntoBayes: An Ontology-Driven Uncertainty Model. In: Proceedings of International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05). Vienna, Austria: IEEE Computer Society; 2005. p.457-63.

7 - Maniraj V, Sivakumar R. Ontology Languages - A Review. Int J Comp Theory and Eng (IJCTE). 2010;2;6:887-91.

8 - Grau BC, Horrocks I, Motik B, Parsia B, Patel-Schenider P, Sattler U. OWL 2: The next step for OWL. Web Semantics: Science, Services and Agents on the World Wide Web. 2008; 6;4: 309-22.

9 - Hitzler P, Krötzsch M, Parsia B, Patel-Schneider PF, Rudolph S. OWL 2 Web Ontology Language Primer (Second Edition). W3C, 2012. http://www.w3.org/TR/2012/REC-owl2-primer-20121211. Accessed 28 Sep 2014.

10 - Aranguren ME, Antezana E, Kuiper M, Stevens R. Ontology Design Patterns Public Catalog. http://odps.sourceforge.net/odp/html (2009). Acessed 28 Sep 2014.

11 - Gangemi A, Presutti V. Ontology Design Patterns. http://ontologydesignpatterns.org (2014). Accessed 28 Sep 2014.

12 - Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean M. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C, 2004. http://www.w3.org/Submission/SWRL. Accessed 28 Sep 2014.

13 - Humphrey SM. Knowledge-based systems for indexing. In: Challenges in indexing electronic text and images. In: Fidel R. et al, editors. Challenges in indexing electronic text and images. Medford, EUA: American Society for Information Science by Learned Information; 1994. p. 161-75.

14 - Joo W, Cha SS. Knowledge-Based Hybrid Expert System for Automated Interferometric Data Reduction. Optics Lasers in Engineering. 1996;24;1-2:15-29.

15 - Hayes-Roth F. Knowledge systems: An introduction. Library Hi Tech. 1992;10;1-2;15-29.

16 – Fogazzi GB, Grignani S. Urinary microscopic analysis: an art abandoned by nephrologists? Nephrology Dialysis Transplantation. 1998;13;10: 2485-7.

17 - Fogazzi GB, Verdesca S, Garigali G. Urianalysis: Core Curriculum 2008. Am J Kidney Dis. 2008;51;6: 1052-67.

18 - Agnew NM, Ford KM, Hayes PJ. Expertise in Context: personally contructed, socially selected, and reality-relevant? Int J Expert Systems. 1994;7;1:65-88.

19 - Turban E. Decision-support and expert systems: management support systems. 3rd ed. Upper Saddle River, USA: Prentice Hall; 1993.

20 - Abel M, CASTILHO JMV. Hybrid information systems: integrating data and knowledge management. In: Proceedings of Conferencia Internacional de la Sociedad Chilena de Ciencia de la Computacion. Santiago, Chile: Sociedad Chilena de Ciencia de la Computacion; 1993. pp. 137-147.

21 - VanLehn K. Problem-solving and cognitive skill acquisition. In: Posner M I, editor. Foundations of Cognitive Science. Cambridge: MIT Press; 1989. p.526-79

22 - Zeitz CM. Some concrete advantages of abstraction: How expert's representation facilitate reasoning. In: Expertise in Context: human and machine. Cambridge: AAAI Press/The MIT Press; 1997. p.43-65.

23 - Abel M. Estudo da perícia em petrografia sedimentar e sua importância para a engenharia de conhecimento. Universidade Federal do Rio Grande do Sul - Programa de Pós-Graduação em Ciência da Computação. PhD thesis, 2001.

24 - Guizzardi G. Ontological foundations for structural conceptual models. Centre for Telematics and Information Technology (CTIT), Enschede. PhD thesis, 2005.

25 - Horridge M, Bechhofer S. The OWL API: A Java API for OWL Ontologies. Sem Web J. 2011;2;1:11-21.

26 - Shearer, R., Motik, B., Horrocks, I. HermiT: A Highly-Efficient OWL Resoner. In: Proceedings of the Fifth International Workshop OWL – Experiences and Directions; 2008.

27 - Rodrigues FH, Flores CD, Rotta LN. An Ontology Design Pattern to Represent Universal Relationships. In: Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015); 2015. p. 17-23.

28 - Guizzardi G, de Almeida Falbo R, Guizzardi RSS. Grounding software domain ontologies in the unified foundational ontology (ufo): The case of the ODE software process ontology. In: Proceedings of CIbSE; 2008. p. 127-40.

CONCLUSÃO

O protótipo desenvolvido neste trabalho obteve desempenho satisfatório na tarefa para a qual foi projetado (i.e. apresentar desempenho comparável ao de um profissional na análise de dados obtidos de amostras de urina). Esse resultado sugere que os modelos ontológicos propostos são adequados para a representação do conhecimento sobre uroanálise necessário durante a realização do exame. Dessa forma, se equipado com interfaces adequadas, além de guiar o processo de análise de amostras, o sistema parece ser adequado para treinamento de profissionais, para consulta e confirmação de hipóteses durante o processo e para automatização parcial da análise.

De toda forma, os resultados obtidos até este momento são preliminares, sendo necessárias novas rodadadas de testes com um conjunto maior e mais variado de casos.

ANEXOS

1 - INSTRUCTIONS FOR AUTHORS (BMC BIOINFORMATICS)

Instructions for authors

Research articles

Criteria | Submission process | Preparing main manuscript text | Preparing illustrations and figures |Preparing tables | Preparing additional files | Style and language

Assistance with the process of manuscript preparation and submission is available from BioMed Central customer support team. See 'About this journal' for information about policies and the refereeing process. We also provide a collection of links to useful tools and resources for scientific authors on our page.

Criteria

Research articles should report on original primary research, but may report on systematic reviews of published research provided they adhere to the appropriate reporting guidelines which are detailed in our Editorial Policies. Please note that non-commissioned pooled analyses of selected published research will not be considered.

Submission process

Manuscripts must be submitted by one of the authors of the manuscript, and should not be submitted by anyone on their behalf. The submitting author takes responsibility for the article during submission and peer review.

Please note that *BMC Bioinformatics* levies an article-processing charge on all accepted Research articles; if the submitting author's institution is a BioMed Central member the cost of the article-processing charge may be covered by the membership (see About page for detail). Please note that the membership is only automatically recognised on submission if the submitting author is based at the member institution.
To facilitate rapid publication and to minimize administrative costs, *BMC Bioinformatics* prefers online submission. Files can be submitted as a batch, or one by one. The submission process can be interrupted at any time; when users return to the site, they can carry on where they left off.

See below for examples of word processor and graphics file formats that can be accepted for the main manuscript document by the online submission system. Additional files of any type, such asmovies, animations, or original data files, can also be submitted as part of the manuscript.
During submission you will be asked to provide a cover letter. Use this to explain why your manuscript should be published in the journal, to elaborate on any issues relating to our editorial policies in the 'About *BMC Bioinformatics*' page, and to declare any potential competing interests. You will be also asked to provide the contact details (including email addresses) of potential peer reviewers for your manuscript. These should be experts in their field, who will be able to provide an objective assessment of the manuscript. Any suggested peer reviewers should not have published with any of the authors of the manuscript within the past five years, should not be current collaborators, and should not be members of the same research institution. Suggested reviewers will be considered alongside potential reviewers recommended by the Editorial team, Editorial Advisors, Section Editors and Associate Editors.
Assistance with the process of manuscript preparation and submission is available from BioMed Central customer support team.
We also provide a collection of links to useful tools and resources for scientific authors on our Useful Tools page.

File formats

The following word processor file formats are acceptable for the main manuscript document:

- Microsoft word (DOC, DOCX)
- Rich text format (RTF)
- Portable document format (PDF)
- TeX/LaTeX (use BioMed Central's TeX template)
- DeVice Independent format (DVI)

TeX/LaTeX users: Please use BioMed Central's TeX template and BibTeX stylefile if you use TeX format. During the TeX submission process, please submit your TeX file as the main manuscript file and your bib/bbl file as a dependent file. Please also convert your TeX file into a PDF and submit this PDF as an additional file with the name 'Reference PDF'. This PDF will be used by internal staff as a reference point to check the layout of the article as the author intended. Please also note that all figures must be coded at the end of the TeX file and not inline.
If you have used another template for your manuscript, or if you do not wish to use BibTeX, then please submit your manuscript as a DVI file. We do not recommend converting to RTF.

For all TeX submissions, all relevant editable source must be submitted during the submission process. Failing to submit these source files will cause unnecessary delays in the publication procedures.

Publishing Datasets

Through a special arrangement with LabArchives, LLC, authors submitting manuscripts to BMC Bioinformatics can obtain a complimentary subscription to LabArchives with an allotment of 100MB of storage. LabArchives is an Electronic Laboratory Notebook which will enable scientists to share and publish data files in situ; you can then link your paper to these data. Data files linked to published articles are assigned digital object identifiers (DOIs) and will remain available in perpetuity. Use of LabArchives or similar data publishing services does not replace preexisting data deposition requirements, such as for nucleic acid sequences, protein sequences and atomic coordinates. Instructions on assigning DOIs to datasets, so they can be permanently linked to publications, can be found on the LabArchives website. Use of LabArchives' software has no influence on the editorial decision to accept or reject a manuscript.

Authors linking datasets to their publications should include an Availability of supporting data section in their manuscript and cite the dataset in their reference list.

Preparing main manuscript text

General guidelines of the journal's style and language are given below.

Overview of manuscript sections for Research articles

Manuscripts for Research articles submitted to *BMC Bioinformatics* should be divided into the following sections (in this order):

- Title page
- Abstract
- Keywords
- Background
- Results and discussion
- Conclusions
- Methods (can also be placed after Background)
- Availability of supporting data
- List of abbreviations used (if any)
- Competing interests
- Authors' contributions

- Authors' information
- Acknowledgements
- Endnotes
- References
- Illustrations and figures (if any)
- Tables and captions
- Preparing additional files

The **Accession Numbers** of any nucleic acid sequences, protein sequences or atomic coordinates cited in the manuscript should be provided, in square brackets and include the corresponding database name; for example, [EMBL:AB026295, EMBL:AC137000, DDBJ:AE000812, GenBank:U49845, PDB:1BFM, Swiss-Prot:Q96KQ7, PIR:S66116].

The databases for which we can provide direct links are: EMBL Nucleotide Sequence Database (EMBL), DNA Data Bank of Japan (DDBJ), GenBank at the NCBI (GenBank), Protein Data Bank (PDB), Protein Information Resource (PIR) and the Swiss-Prot Protein Database (Swiss-Prot).

You can download a template (Mac and Windows compatible; Microsoft Word 98/2000) for your article.

For reporting standards please see the information in the About section.

Title page

The title page should:

- provide the title of the article
- list the full names, institutional addresses and email addresses for all authors
- indicate the corresponding author

Please note:

- abbreviations within the title should be avoided

Abstract

The Abstract of the manuscript should not exceed 350 words and must be structured into separate sections: **Background,** the context and purpose of the study; **Results**, the main findings;**Conclusions**, brief summary and potential implications. Please minimize the use of abbreviations and do not cite references in the abstract.

Keywords

Three to ten keywords representing the main content of the article.

Background

The Background section should be written in a way that is accessible to researchers without specialist knowledge in that area and must clearly state - and, if helpful, illustrate - the background to the research and its aims. The section should end with a brief statement of what is being reported in the article.

Results and discussion

The Results and discussion may be combined into a single section or presented separately. The Results and discussion sections may also be broken into subsections with short, informative headings.

Conclusions

This should state clearly the main conclusions of the research and give a clear explanation of their importance and relevance. Summary illustrations may be included.

Methods

The methods section should include the design of the study, the type of materials involved, a clear description of all comparisons, and the type of analysis used, to enable replication. For studies involving human participants a statement detailing ethical approval and consent should be included in the methods section. For further details of the journal's editorial policies and ethical guidelines see 'About this journal'.

For further details of the journal's data-release policy, see the policy section in 'About this journal'.

Availability of supporting data

BMC Bioinformatics encourages authors to deposit the data set(s) supporting the results reported in submitted manuscripts in a publicly-accessible data repository, when it is not possible to publish them as additional files. This section should only be included when supporting data are available and must include the name of the repository and the permanent identifier or accession number and persistent hyperlink(s) for the data set(s). The following format is required:
"The data set(s) supporting the results of this article is(are) available in the [repository name] repository, [unique persistent identifier and hyperlink to dataset(s) in http:// format]."

Where all supporting data are included in the article or additional files the following format is required:

"The data set(s) supporting the results of this article is(are) included within the article (and its additional file(s))"

We also recommend that the data set(s) be cited, where appropriate in the manuscript, and included in the reference list.

A list of available scientific research data repositories can be found here. A list of all BioMed Central journals that require or encourage this section to be included in research articles can be found here.

List of abbreviations

If abbreviations are used in the text they should be defined in the text at first use, and a list of abbreviations can be provided, which should precede the competing interests and authors' contributions.

Competing interests

A competing interest exists when your interpretation of data or presentation of information may be influenced by your personal or financial relationship with other people or organizations. Authors must disclose any financial competing interests; they should also reveal any non-financial competing interests that may cause them embarrassment were they to become public after the publication of the manuscript.

Authors are required to complete a declaration of competing interests. All competing interests that are declared will be listed at the end of published articles. Where an author gives no competing interests, the listing will read 'The author(s) declare that they have no competing interests'.

When completing your declaration, please consider the following questions:

Financial competing interests

- In the past three years have you received reimbursements, fees, funding, or salary from an organization that may in any way gain or lose financially from the publication of this manuscript, either now or in the future? Is such an organization financing this manuscript (including the article-processing charge)? If so, please specify.
- Do you hold any stocks or shares in an organization that may in any way gain or lose financially from the publication of this manuscript, either now or in the future? If so, please specify.
- Do you hold or are you currently applying for any patents relating to the content of the manuscript? Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript? If so, please specify.
- Do you have any other financial competing interests? If so, please specify.

Non-financial competing interests

Are there any non-financial competing interests (political, personal, religious, ideological, academic, intellectual, commercial or any other) to declare in relation to this manuscript? If so, please specify.

If you are unsure as to whether you, or one your co-authors, has a competing interest please discuss it with the editorial office.

Authors' contributions

In order to give appropriate credit to each author of a paper, the individual contributions of authors to the manuscript should be specified in this section.

According to ICMJE guidelines, An 'author' is generally considered to be someone who has made substantive intellectual contributions to a published study. To qualify as an author one should 1) have made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) have been involved in drafting the manuscript or revising it critically for important intellectual content; 3) have given final approval of the version to be published; and 4) agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content. Acquisition of funding, collection of data, or general supervision of the research group, alone, does not justify authorship.
We suggest the following kind of format (please use initials to refer to each author's contribution): AB carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. JY carried out the immunoassays. MT participated in the sequence alignment. ES participated in the design of the study and performed the statistical analysis. FG conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

All contributors who do not meet the criteria for authorship should be listed in an acknowledgements section. Examples of those who might be acknowledged include a person who provided purely technical help, writing assistance, or a department chair who provided only general support.

Authors' information

You may choose to use this section to include any relevant information about the author(s) that may aid the reader's interpretation of the article, and understand the standpoint of the author(s). This may include details about the authors' qualifications, current positions they hold at institutions or societies, or any other relevant background information. Please refer to authors using their initials. Note this section should not be used to describe any competing interests.

Acknowledgements

Please acknowledge anyone who contributed towards the article by making substantial contributions to conception, design, acquisition of data, or analysis and interpretation of data, or who was involved in drafting the manuscript or revising it critically for important intellectual content, but who does not meet the criteria for authorship. Please also include the source(s) of funding for each author, and for the manuscript preparation. Authors must describe the role of the funding body, if any, in design, in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication. Please also acknowledge anyone who contributed materials essential for the study. If a language editor has made significant revision of the manuscript, we recommend that you acknowledge the editor by name, where possible.

Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section.

Endnotes

Endnotes should be designated within the text using a superscript lowercase letter and all notes (along with their corresponding letter) should be included in the Endnotes section. Please format this section in a paragraph rather than a list.

References

All references, including URLs, must be numbered consecutively, in square brackets, in the order in which they are cited in the text, followed by any in tables or legends. Each reference must have an individual reference number. Please avoid excessive referencing. If automatic numbering systems are used, the reference numbers must be finalized and the bibliography must be fully formatted before submission.

Only articles and abstracts that have been published or are in press, or are available through public e-print/preprint servers, may be cited; unpublished abstracts, unpublished data and personal communications should not be included in the reference list, but may be included in the text and referred to as "unpublished observations" or "personal communications" giving the names of the involved researchers. Obtaining permission to quote personal communications and unpublished data from the cited colleagues is the responsibility of the author. Footnotes are not allowed, but endnotes are permitted. Journal abbreviations follow Index Medicus/MEDLINE. Citations in the reference list should include all named authors, up to the first 30 before adding '*et al.*'..

Any *in press* articles cited within the references and necessary for the reviewers' assessment of the manuscript should be made available if requested by the editorial office.

An Endnote style file is available.

Examples of the *BMC Bioinformatics* reference style are shown below. Please ensure that the reference style is followed precisely; if the references are not in the correct style they may have to be retyped and carefully proofread.

All web links and URLs, including links to the authors' own websites, should be given a reference number and included in the reference list rather than within the text of the manuscript. They should be provided in full, including both the title of the site and the URL, in the following format: **The Mouse Tumor Biology Database** [http://tumor.informatics.jax.org/mtbwi/index.do]. If an author or group of authors can clearly be associated with a web link, such as for weblogs, then they should be included in the reference.

Examples of the *BMC Bioinformatics* reference style

Article within a journal

Smith JJ. The world of science. Am J Sci. 1999;36:234-5.

Article by DOI (with page numbers)

O'Mahony S, Rose SL, Chilvers AJ, Ballinger JR, Solanki CK, Barber RW. Finding an optimal method for imaging lymphatic vessels of the upper limb. Eur J Nucl Med Mol Imaging. 2004;31:555-63. doi:10.1007/s00259-003-1399-3.

Article by DOI (before issue publication and without page numbers)

O'Mahony S, Rose SL, Chilvers AJ, Ballinger JR, Solanki CK, Barber RW. Finding an optimal method for imaging lymphatic vessels of the upper limb. Eur J Nucl Med Mol Imaging. 2004. doi:10.1007/s00259-003-1399-3.

Article in electronic journal by DOI (no paginated version)

Slifka MK, Whitton JL. Clinical implications of dysregulated cytokine production. Dig J Mol Med. 2000. doi:10.1007/s801090000086.

Article within a journal supplement

Frumin AM, Nussbaum J, Esposito M. Functional asplenia: demonstration of splenic activity by bone marrow scan. Blood 1979;59 Suppl 1:26-32.

Book chapter, or an article within a book

Wyllie AH, Kerr JFR, Currie AR. Cell death: the significance of apoptosis. In: Bourne GH,

Danielli JF, Jeon KW, editors. International review of cytology. London: Academic; 1980. p. 251-306.

OnlineFirst chapter in a series (without a volume designation but with a DOI)

Saito Y, Hyuga H. Rate equation approaches to amplification of enantiomeric excess and chiral symmetry breaking. Top Curr Chem. 2007. doi:10.1007/128_2006_108.

Complete book, authored

Blenkinsopp A, Paxton P. Symptoms in the pharmacy: a guide to the management of common illness. 3rd ed. Oxford: Blackwell Science; 1998.

Online document

Doe J. Title of subordinate document. In: The dictionary of substances and their effects. Royal Society of Chemistry. 1999. http://www.rsc.org/dose/title of subordinate document. Accessed 15 Jan 1999.

Online database

Healthwise Knowledgebase. US Pharmacopeia, Rockville. 1998. http://www.healthwise.org. Accessed 21 Sept 1998.

Supplementary material/private homepage

Doe J. Title of supplementary material. 2000. http://www.privatehomepage.com. Accessed 22 Feb 2000.

University site

Doe, J: Title of preprint. http://www.uni-heidelberg.de/mydata.html (1999). Accessed 25 Dec 1999.

FTP site

Doe, J: Trivial HTTP, RFC2169. ftp://ftp.isi.edu/in-notes/rfc2169.txt (1999). Accessed 12 Nov 1999.

Organization site

ISSN International Centre: The ISSN register. http://www.issn.org (2006). Accessed 20 Feb 2007.

Preparing illustrations and figures

Illustrations should be provided as separate files, not embedded in the text file. Each figure should include a single illustration and should fit on a single page in portrait format. If a figure consists of separate parts, it is important that a single composite illustration file be submitted which contains all parts of the figure. There is no charge for the use of color figures.

Please read our figure preparation guidelines for detailed instructions on maximising the quality of your figures.

Formats

The following file formats can be accepted:

- PDF (preferred format for diagrams)
- DOCX/DOC (single page only)
- PPTX/PPT (single slide only)
- EPS
- PNG (preferred format for photos or images)
- TIFF
- JPEG
- BMP

Figure legends

The legends should be included in the main manuscript text file at the end of the document, rather than being a part of the figure file. For each figure, the following information should be provided: Figure number (in sequence, using Arabic numerals - i.e. Figure 1, 2, 3 etc); short title of figure (maximum 15 words); detailed legend, up to 300 words.

Please note that it is the responsibility of the author(s) to obtain permission from the copyright holder to reproduce figures or tables that have previously been published elsewhere.

Preparing tables

Each table should be numbered and cited in sequence using Arabic numerals (i.e. Table 1, 2, 3 etc.). Tables should also have a title (above the table) that summarizes the whole table; it should be no longer than 15 words. Detailed legends may then follow, but they should be concise. Tables should always be cited in text in consecutive numerical order.

Smaller tables considered to be integral to the manuscript can be pasted into the end of the document text file, in A4 portrait or landscape format. These will be typeset and displayed in the final published form of the article. Such tables should be formatted using the 'Table object' in a word processing program to ensure that columns of data are kept aligned when the file is sent electronically for review; this will not always be the case if columns are generated by simply using tabs to separate text. Columns and rows of data should be made visibly distinct by ensuring that the borders of each cell display as black lines. Commas should not be used to indicate numerical values. Color and shading may not be used; parts of the table can be highlighted using symbols or bold text, the meaning of which should be explained in a table legend. Tables should not be embedded as figures or spreadsheet files.

Larger datasets or tables too wide for a portrait page can be uploaded separately as additional files. Additional files will not be displayed in the final, laid-out PDF of the article, but a link will be provided to the files as supplied by the author.

Tabular data provided as additional files can be uploaded as an Excel spreadsheet (.xls) or comma separated values (.csv). As with all files, please use the standard file extensions.

Preparing additional files

Although *BMC Bioinformatics* does not restrict the length and quantity of data included in an article, we encourage authors to provide datasets, tables, movies, or other information as additional files.
Please note: All Additional files **will be published** along with the article. Do not include files such as patient consent forms, certificates of language editing, or revised versions of the main manuscript document with tracked changes. Such files should be sent by email to editorial@biomedcentral.com, quoting the Manuscript ID number.
Results that would otherwise be indicated as "data not shown" can and should be included as additional files. Since many weblinks and URLs rapidly become broken, *BMC Bioinformatics* requires that supporting data are included as additional files, or deposited in a recognized repository. Please do not link to data on a personal/departmental website. The maximum file size for additional files is 20 MB each, and files will be virus-scanned on submission.
Additional files can be in any format, and will be downloadable from the final published article as supplied by the author. We recommend CSV rather than PDF for tabular data.

Certain supported files formats are recognized and can be displayed to the user in the browser. These include most movie formats (for users with the Quicktime plugin), mini-websites prepared according to our guidelines, chemical structure files (MOL, PDB), geographic data files (KML).

If additional material is provided, please list the following information in a separate section of the manuscript text:

- File name (e.g. Additional file 1)
- File format including the correct file extension for example .pdf, .xls, .txt, .pptx (including name and a URL of an appropriate viewer if format is unusual)
- Title of data
- Description of data

Additional files should be named "Additional file 1" and so on and should be referenced explicitly by file name within the body of the article, e.g. 'An additional movie file shows this in more detail [see Additional file 1]'.

Additional file formats

Ideally, file formats for additional files should not be platform-specific, and should be viewable using free or widely available tools. The following are examples of suitable formats.

- Additional documentation
  - PDF (Adode Acrobat)
- Animations
  - SWF (Shockwave Flash)
- Movies
  - MP4 (MPEG 4)
  - MOV (Quicktime)
- Tabular data
  - XLS, XLSX (Excel Spreadsheet)
  - CSV (Comma separated values)

As with figure files, files should be given the standard file extensions.

Mini-websites

Small self-contained websites can be submitted as additional files, in such a way that they will be browsable from within the full text HTML version of the article. In order to do this, please follow these instructions:

1. Create a folder containing a starting file called index.html (or index.htm) in the root.
2. Put all files necessary for viewing the mini-website within the folder, or sub-folders.
3. Ensure that all links are relative (ie "images/picture.jpg" rather than "/images/picture.jpg" or "http://yourdomain.net/images/picture.jpg" or "C:\Documents and Settings\username\My Documents\mini-website\images\picture.jpg") and no link is longer than 255 characters.
4. Access the index.html file and browse around the mini-website, to ensure that the most commonly used browsers (Internet Explorer and Firefox) are able to view all parts of the mini-website without problems, it is ideal to check this on a different machine.
5. Compress the folder into a ZIP, check the file size is under 20 MB, ensure that index.html is in the root of the ZIP, and that the file has .zip extension, then submit as an additional file with your article.

Style and language

General

Currently, *BMC Bioinformatics* can only accept manuscripts written in English. Spelling should be US English or British English, but not a mixture.
There is no explicit limit on the length of articles submitted, but authors are encouraged to be concise.

BMC Bioinformatics will not edit submitted manuscripts for style or language; reviewers may advise rejection of a manuscript if it is compromised by grammatical errors. Authors are advised to write clearly and simply, and to have their article checked by colleagues before submission. In-house copyediting will be minimal. Non-native speakers of English may choose to make use of a copyediting service.

Language editing

For authors who wish to have the language in their manuscript edited by a native-English speaker with scientific expertise, BioMed Central recommends Edanz. BioMed Central has arranged a 10% discount to the fee charged to BioMed Central authors by Edanz. Use of an editing service is neither a requirement nor a guarantee of acceptance for publication. Please contact Edanz directly to make arrangements for editing, and for pricing and payment details.

Help and advice on scientific writing

The abstract is one of the most important parts of a manuscript. For guidance, please visit our page on Writing titles and abstracts for scientific articles.
Tim Albert has produced for BioMed Central a list of tips for writing a scientific manuscript. American Scientist also provides a list of resources for science writing. For more detailed guidance on preparing a manuscript and writing in English, please visit the BioMed Central author academy.

Abbreviations

Abbreviations should be used as sparingly as possible. They should be defined when first used and a list of abbreviations can be provided following the main manuscript text.

Typography

- Please use double line spacing.
- Type the text unjustified, without hyphenating words at line breaks.
- Use hard returns only to end headings and paragraphs, not to rearrange lines.
- Capitalize only the first word, and proper nouns, in the title.
- All lines and pages should be numbered. Authors are asked to ensure that line numbering is included in the main text file of their manuscript at the time of submission to facilitate peer-review. Once a manuscript has been accepted, line numbering should be removed from the manuscript before publication. For authors submitting their manuscript in Microsoft Word please do not insert page breaks in your manuscript to ensure page numbering is consistent between your text file and the PDF generated from your submission and used in the review process.
- Use the *BMC Bioinformatics* reference format.
- Footnotes are not allowed, but endnotes are permitted.
- Please do not format the text in multiple columns.
- Greek and other special characters may be included. If you are unable to reproduce a particular special character, please type out the name of the symbol in full. **Please ensure that all special characters used are embedded in the text, otherwise they will be lost during conversion to PDF.**
- Genes, mutations, genotypes, and alleles should be indicated in italics, and authors are required to use approved gene symbols, names, and formatting. Protein products should be in plain type.

Units

SI units should be used throughout (liter and molar are permitted, however).

2 - REGISTRO DO PROJETO NA COMISSÃO DE PESQUISA

REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DA EDUCAÇÃO

UFCSPA
UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE PORTO ALEGRE
COMISSÃO DE PESQUISA

Atestado

Atestamos, para o fim de inscrição de projeto de pesquisa no processo seletivo de iniciação científica e iniciação tecnológica e inovação da UFCSPA, que os projetos de pesquisa abaixo listados estão registrados na Comissão de Pesquisa:

Pesquisador Responsável	Título do projeto de pesquisa
Ana Beatriz Gorini da Veiga	Avaliação epidemiológica do perfil de resistência dos vírus da hepatite B e C: identificação de sorotipos, subgenótipos e mutações virais de resistência primária aos antivirais e caracterização de genes humanos associados à resposta virológica sustentada e ao prognóstico em pacientes infectados e não infectados pelo HIV
Ana Cristina Borba da Cunha	Aplicação de processos de biotransformação para degradação de benzo[a]pireno em solos contaminados com benzo[a]pireno
Cícero Armídio Gomes Dias	Prevalência de resistência aos antimicrobianos e diversidade genética de Streptococcus pneumoniae isolados de doença invasiva
Dinara Jaqueline Moura	A proteína Kin3 de Saccharomyces cerevisiae: estudo estrutural e funcional
Eliane Dallegrave	Avaliação da atividade antibacteriana, antifúngica, antiviral e antitumoral do veneno de serpentes do gênero Bothrops
Jenifer Saffi	Bioprospecção sistemática de extratos vegetais para a busca de novos agentes antineoplásicos
Jenifer Saffi	Prospecção de novos compostos sintéticos pirimídicos como potenciais medicamentos para a terapia do câncer
Josiane Somariva Prophiro	Estudo da atividade tripanocida e imunomoduladora de Photorhabdus luminescens e Xenorhabdus nematophilia (Enterobacteriaceae)
Liane Nanci Rotta	Sistema baseado em conhecimento para uroanálise
Márcia Vignoli da Silva	Caracterização morfoanatômica e histoquímica das folhas de Psidium cattleianum Sabine (Myrtaceae)
Márcia Vignoli da Silva	Caracterização morfoanatômica e histoquímica de espécies do gênero Opuntia Mill. (CACTACEAE) no Rio Grande do Sul
Pedro Roosevelt Torres Romão	Investigação da atividade imunomoduladora e anti-leishmania de espécies do gênero Hypericum
Pedro Roosevelt Torres Romão	Investigação da atividade leishmanicida e imunomoduladora de produtos de bactérias entomopatogênicas

Porto Alegre, 07 de maio de 2014.

Paulo Ricardo Gazzola Zen
Coordenador Geral da Pesquisa
UFCSPA

Rua Sarmento Leite, 245 • Porto Alegre, RS • CEP 90050-170 • Fone: 0 xx 51 3303 9000 • Fax: 0 xx 51 3303.8810 • www.ufcspa.edu.br

3 - MATERIAIS CIENTÍFICOS PUBLICADOS/ACEITOS PARA PUBLICAÇÃO DURANTE O PERÍODO DE REALIZAÇÃO DA DISSERTAÇÃO (2013-2015).

1 - Rodrigues, F. H., Bez, M. R., Flores, C. D. Generating Bayesian networks from medical ontologies. In: Proceedings of Computing Colombian Conference (8CCC), 2013.

2 - Rodrigues, F. H., Flores, C. D., Bez, M. R. Geração de Redes Bayesianas para Clinica Médica a partir de Ontologias Médicas. In: Anais do 12º Congresso Brasileiro de Clinica Médica, 2013.

3 - Rodrigues, F. H., Poloni, J. A. T., Flores, C. D., Rotta, L. N. Knowledge-based System for Urinalysis. In: Proceedings of 16th International Conference on Enterprise Information Systems (ICEIS), 2014.

4 - Rodrigues, F. H., Flores, C. D., Rotta, L. N. An Ontology Design Pattern to Represent Universal Relationships. ICSC 2015 (aceito para publicação).