

# Utilizando XML para representação de informação em Saúde

**Fabiane Bizinella Nardon**  
**Unidade de Pesquisa e Desenvolvimento**  
**Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da USP**  
**São Paulo – SP**  
fabiane.nardon@incor.usp.br

Prover atendimento de boa qualidade para um paciente depende muito da quantidade e da qualidade da informação que está disponível. Há muito tempo que se sabe disso e a utilização da informática em saúde surgiu justamente com o propósito de facilitar o acesso à informação, auxiliando, assim, a prática médica. A informação em saúde, no entanto, é extremamente complexa e pouco estruturada e durante muito tempo a melhor forma que se encontrou para representar a informação foi através do simples armazenamento de documentos médicos descritos por texto livre. O texto livre é uma ótima forma de representação de informação: com algumas frases, pode-se dizer qualquer coisa, pode-se descrever sintomas, impressões, resultados de exames, enfim, é uma forma de representação extremamente poderosa, facilmente legível por seres humanos e que seria perfeita, não fosse um pequeno detalhe: os computadores são máquinas limitadas demais para compreenderem um documento médico escrito em texto livre! Os computadores precisam de informação estruturada para poder processar eficientemente a informação. Hoje não basta mais ter simplesmente uma grande base de documentos médicos. É importante também, a partir de um conjunto de documentos, ser capaz de perguntar “Qual foi o diagnóstico final?”, “Que sintomas o paciente apresentava quando foi internado?”, “Que outros pacientes apresentaram os mesmos sintomas?”. É oferecendo respostas a estas perguntas que os computadores podem efetivamente auxiliar na prática médica. Em uma base composta por documentos representados como texto livre e não estruturados, no entanto, é praticamente impossível construir um sistema que seja capaz de responder a estas perguntas. O que se precisa é uma forma de representação que seja legível por seres humanos e ao mesmo tempo processável por um computador.

Existem ainda outros requisitos que devem ser considerados no armazenamento de documentos médicos. Como a informação em saúde deve ser armazenada por muito tempo (teoricamente por toda a vida do paciente), é importante que a informação seja armazenada em uma forma capaz de resistir a mudanças tecnológicas. Não adianta armazenar hoje os documentos no formato do editor de textos mais utilizado pela instituição. É muito pouco provável que daqui 30 ou 40 anos este formato de arquivo ainda seja tão popular quanto é hoje e não se pode garantir que a versão do mesmo editor de textos que se estará usando então será capaz de ler arquivos no formato em que eles foram gravados hoje. Além disso, cada vez mais é importante que uma instituição seja capaz de trocar informações com outras instituições e para que isto seja possível é necessário utilizar uma forma de representação de documentos médicos que seja padronizada e independente de plataforma de hardware e software. Adicionalmente, as informações contidas nos documentos médicos deveriam poder ser apresentadas de forma diferente em diferentes aplicações, dependendo da necessidade do usuário. Finalmente, uma vez que tem se tornado cada vez mais claro que o acesso à informação de dará invariavelmente através da internet, deve-se considerar uma representação da informação que seja compatível com a web.

O XML (eXtensible Markup Language) é um padrão para representação de documentos em computador criado pelo W3 Consortium (o mesmo que criou a linguagem HTML) que desde o seu surgimento tem despertado bastante interesse na área de informática em saúde. O princípio do XML é, na verdade, muito simples: qualquer um pode inventar marcações que adicionam semântica a um texto, dizendo o que significa um determinado pedaço de informação. Por exemplo, supondo que se queira armazenar o seguinte pedaço de informação:

“O paciente João da Silva (matrícula número 123) apresenta uma lesão de 40% no 1/3 proximal na Artéria 2a. Marginal Esquerda”

Enquanto o texto acima é perfeitamente compreensível por um ser humano, como poderia-se, por exemplo, construir um sistema que respondesse qual o grau da lesão apresentada pelo paciente, ou quais os pacientes que apresentaram lesão na Artéria 2a. Marginal Esquerda maior que 30%? Uma representação do mesmo texto em XML poderia ser algo como:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Laudo>
  <Paciente matrícula="123">
    <Nome> João da Silva </Nome>
  </Paciente>
  <Conclusões>
    <LesãoArterial>
      <Grau> 40% </Grau>
      <Localização> 1/3 proximal </Localização>
      <Artéria> 2a. Marginal </Artéria>
    </LesãoArterial>
  </Conclusões>
</Laudo>
```

No exemplo acima, a primeira linha (<?xml version="1.0" encoding="ISO-8859-1"?>), indica que o que se está criando um documento XML. O restante do texto é uma mistura de informações sobre o laudo do paciente com marcações que indicam qual a semântica de cada um dos pedaços de informação. Por exemplo, onde se escreve o nome do paciente, existe uma marcação informando que aquela informação é o <Nome>. Estas marcações são chamadas de tags e podem ser criadas livremente pela pessoa que cria o documento. Esta é uma das principais diferenças entre o XML e o HTML. Enquanto no HTML as tags são pré-fixadas, no XML elas podem ser criadas a vontade. Outra grande diferença, é que as tags do HTML especificam formatação e não conteúdo. Por exemplo, a tag <b> do HTML indica que o texto que virá a seguir deve ser apresentado em negrito. No XML, ao contrário, não existe informação sobre como a informação deverá ser apresentada, mas sim sobre o que a informação significa.

Toda tag XML é escrita entre os símbolos < e >. Quando o texto que compõe uma tag termina, é acrescentado um indicativo de fechamento, que é a tag iniciando com o os símbolos </ . Por exemplo, a informação sobre o nome do paciente inicia com a tag <Nome> e termina com a tag </Nome>.

As tags XML podem ser aninhadas em diversos níveis e podem se repetir várias vezes dentro de um mesmo documento. Isto permite que as informações sejam representada em estruturas hierárquicas, adicionando, assim, novos níveis semânticos. Existem dois tipos de informação que podem aparecer em um documento XML: elementos e atributos. Elementos são as informações que são representadas entre as tags, como <Paciente>, <Nome> e <LesãoArterial>, por exemplo. Atributos são informações adicionais que são acrescentadas a uma tag, como o atributo "matrícula", no exemplo anterior. Representar uma informação como elemento ou como atributo depende das características da informação que se está representando. Por exemplo, atributos permitem que se estabeleça um conjunto de valores possíveis, enquanto elementos não permitem isso. Tanto atributos quanto elementos tem suas vantagens e desvantagens e a escolha entre um e outro deve ser feita cuidadosamente.

Observando este primeiro exemplo, já é possível perceber como o XML resolve os diversos problemas discutidos no início deste artigo. Em primeiro lugar, a informação é representada de uma forma que os computadores são capazes de processar. Com a inclusão das tags, é possível criar facilmente um programa que seja capaz de determinar, por exemplo, qual o grau da lesão arterial que o paciente possuía. Um documento no formato XML é um arquivo ASCII simples, o

que significa que ele pode ser lido em qualquer plataforma de hardware e software, permitindo a troca de informações entre sistemas heterogêneos. Como um documento XML é também independente de tecnologia, o que significa que não é necessário nenhum software específico para ler um arquivo XML, os documentos criados neste padrão irão sobreviver a mudanças tecnológicas e o mesmo documento poderá ser visto daqui 30 ou 40 anos. O XML é um padrão que já nasceu para ser utilizado na internet. Alguns browsers já possuem suporte ao XML nativamente, como Internet Explorer 5 e o Mozilla.

Um documento XML pode ser criado livremente, com tags quaisquer, mas na maioria dos casos, é muito importante ter um modelo que defina que tags podem ser usadas em um determinado tipo de documento e como elas podem ser utilizadas. Isto é feito através de um outro padrão criado pelo W3 Consortium chamado Document Type Definition (DTD). Uma DTD descreve a estrutura do documento e, uma vez associada a um documento XML, é utilizada para validá-lo, garantindo que as tags estão sendo utilizadas de forma correta. A criação de uma DTD é feita através da descrição dos elementos e atributos que podem aparecer no documento XML, informando também se os elementos podem se repetir, que valores possíveis um atributo pode assumir e que elementos e atributos são opcionais. Para o documento XML apresentado como exemplo, uma possível DTD seria:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!ELEMENT Laudo (Paciente, Conclusões+)>
<!ELEMENT Paciente (Nome)>
<!ATTLIST Paciente matrícula CDATA #IMPLIED>
<!ELEMENT Nome (#PCDATA)>
<!ELEMENT Conclusões (LesãoArterial?)>
<!ELEMENT LesãoArterial (Grau, Localização, Artéria)>
<!ELEMENT Grau (#PCDATA)>
<!ELEMENT Localização (#PCDATA)>
<!ELEMENT Artéria (#PCDATA)>
```

A primeira linha indica que o documento é um documento XML e a seguir inicia-se a descrição dos elementos e atributos. O elemento Laudo, por exemplo, é composto de um elemento Paciente, cuja presença é obrigatória, e de um ou mais elementos Conclusões. A possibilidade de haver um ou mais elementos Conclusões é indicado pelo sinal + colocado logo após o nome do elemento. O elemento Paciente, por sua vez, é composto de um elemento Nome e de um atributo matrícula. O atributo matrícula conterá um texto (que é o que a palavra CDATA representa) e será opcional (que é o que #IMPLIED representa). O elemento Nome será um texto qualquer e é isso que #PCDATA significa. O elemento Conclusões possui um elemento LesãoArterial, que é opcional. Esta opcionalidade é indicada pelo símbolo ? colocado após o elemento.

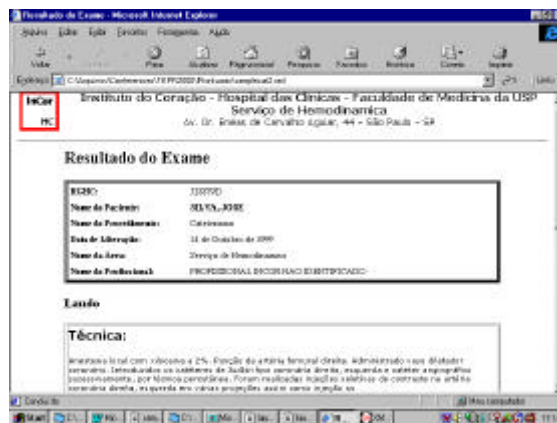
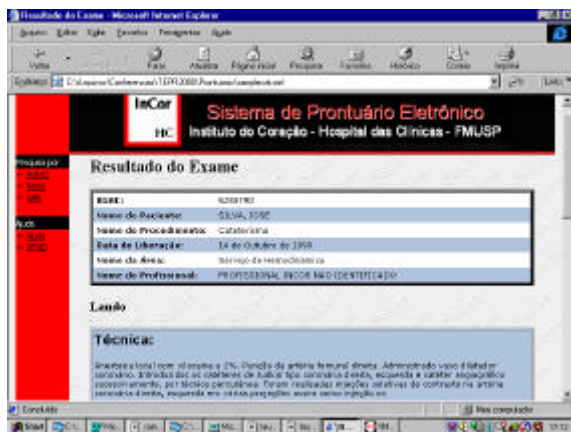
Depois de criada a DTD, basta informar no documento XML que ela deve ser usada para validar o documento. Isto é feito acrescentando a linha <!DOCTYPE Laudo SYSTEM "laudo.dtd"> logo após a linha <?xml version="1.0" encoding="ISO-8859-1"?> no documento XML. Neste exemplo, laudo.dtd é o nome do arquivo que contém a especificação da DTD.

Embora a criação da DTD exija um pouco de familiaridade com sua sintaxe, é muito fácil criar um modelo de documento. Diversas ferramentas capazes de verificar se um documento XML respeita a sua DTD estão disponíveis, como o XML Spy, o XML Authority e muitos outros.

A DTD é a forma com que duas instituições podem estabelecer o formato comum em que irão trocar informações. Se houverem DTDs padronizadas, será possível trocar laudos de exames, por exemplo. Um esforço de padronização que já existe na área da saúde é o Comitê X31.25 do ASTM, cujo objetivo é criar um conjunto de DTDs padronizadas para representação de documentos médicos.

Embora bastante útil, a DTD ainda possui vários pontos em que deixa a desejar. Um destes pontos é que todos os elementos descritos em uma DTD são do tipo texto. Não é possível em uma DTD especificar que um determinado elemento só poderá conter números, ou só datas válidas. Em função desta limitação, o W3 Consortium está criando uma nova forma de representar modelos de documentos. Esta nova forma se chama XML Schema. O Schema é muito mais poderoso que a DTD, mas sua especificação ainda não está concluída e portanto tem-se que esperar mais algum tempo até que se possa utilizá-la. Já existem, no entanto, algumas ferramentas que suportam o XML Schema, embora tenham sido construídas a partir de drafts do padrão.

Tendo criado o arquivo XML e a DTD, falta apenas descobrir como apresentar o XML para o usuário. Uma das maiores vantagens do XML é justamente a possibilidade de separar conteúdo de apresentação. Uma vez que o XML não possui nenhuma indicação de como a informação será apresentada, é possível apresentar o mesmo documento de formas diferentes para usuários diferentes, ou em aplicações diferentes. As figuras abaixo mostram duas visões de um mesmo laudo de cateterismo criados pelo Sistema de Prontuário Eletrônico do Instituto do Coração. A primeira visão é a que normalmente um usuário veria quando estivesse usando o sistema via Web. A segunda visão apresenta uma formatação do laudo própria para ser impressa, onde banners e cores são retiradas para otimizar a impressão.



A apresentação de um documento XML depende da folha de estilos (stylesheet) que está associada ao documento. Os dois padrões mais utilizados para construir folhas de estilos são o CSS (Cascading Style Sheet) e o XSL (Extensible Style Language). O CSS é suportado por quase todos os browsers, mas é muito pobre em termos do que é capaz de construir. Não se pode com um CSS, por exemplo, mostrar os elementos em outra ordem que não a apresentada no documento, nem não apresentar um determinado elemento. Já o XSL é extremamente poderoso e é capaz de transformar um documento XML em HTML, PDF, RTF e uma série de outros formatos. O maior problema do XSL é que a sua especificação final ainda não foi concluída pelo W3 Consortium e por isso poucas ferramentas suportam este formato. O Internet Explorer 5 implementa um dos últimos drafts desta especificação e o que se pode fazer com ela é realmente impressionante. O prontuário eletrônico do Instituto do Coração usa o XSL para fazer a transformação dos documentos XML para HTML e muitas vantagens advêm desta tecnologia. Com o XSL é possível, por exemplo, mostrar os valores anormais de um exame com outra cor, ou permitir que o usuário solicite dinamicamente uma ordenação diferente das informações que estão sendo apresentadas.

Sem dúvida nenhuma, o maior problema do XML atualmente é a falta de ferramentas que auxiliem a criação das folhas de estilo. Muito do trabalho ainda deve ser feito manualmente, mas isto é, sem dúvida nenhuma, uma questão de tempo.

O que realmente é importante na área da saúde é ter-se modelos de documentos XML padronizados. Isto nos permitirá trocar informações entre as instituições, melhorando muito a qualidade da informação para o paciente. O Comitê de Padronização do Registro Clínico já está preparando uma versão da sua especificação de dados mínimos do paciente em XML. Além do esforço de padronização do ASTM, já citado anteriormente, o HL7 está criando documentos XML que representarão as suas mensagens e este trabalho já está bastante avançado. O CORBAMED também tem usado o XML em algumas de suas especificações. O Serviço de Acesso a Observações Clínicas, por exemplo, suporta a troca de observações no formato XML. O padrão DICOM também está considerando a representação dos laudos que acompanham as imagens em XML.

Enfim, o XML representa uma forma de representação de documentos muito interessante, principalmente para a área da saúde. O XML pode ser a solução para representar documentos com o poder expressivo do texto livre e com a estrutura necessária para o processamento computadorizado da informação.

Onde saber mais:

W3 Consortium: <http://www.w3.org/>

XML.ORG: [www.xml.org](http://www.xml.org)

HL7 XML Special Interest Group: <http://www.hl7.org/Special/committees/sgml/sgml.htm>