

Produção de Notícias Multimídia para Diferentes Meios de Comunicação

Cássio V. S. Prazeres¹, Celso A. S. Santos¹, Cesar A. C. Teixeira²

¹Universidade Salvador – Núcleo Interdepartamental de Pesquisa em Redes de Computadores (NUPERC) – Salvador-BA

²Faculdades COC – Ribeirão Preto-SP

cassio@unifacs.br, saibel@unifacs.br, cesar.teixeira@coc.com.br

Abstract. *The advent of the Web as a new publishing media where news may be presented as a multimedia and interactive document and the digitalization of content are imposing changes in the current models for producing journalistic content. These changes look for a rationalization in the way journalists compose their material (possibly a multimedia document), mainly when the same news must be presented through different communication media and be available for retrieving by tools oriented by metadata, markup and hyperlinks established over contents. The aims of this paper are to present and to discuss some problems and solutions related to the authoring of hypermedia/multimedia news. Issues like multimedia news content production (Multimedia Journalism), news content production for presenting in multiple media of communication, preserving semantic of news content and establishing hyperlinks among correlated news are discussed in the paper. A model to represent and manage news, based on markup languages, is presented, together with a tool designed to minimize possible resistance journalists may offer to these changes.*

Resumo. *O advento da Web como um novo meio de publicação onde a notícia pode ser apresentada como um documento multimídia e interativo e a digitalização do conteúdo estão impondo mudanças nos modelos atuais para produção de conteúdos jornalísticos. Estas mudanças buscam racionalizar a maneira com que os jornalistas compõem seu material (possivelmente um documento multimídia), principalmente quando a mesma notícia deve ser apresentada em diferentes meios de comunicação e estar disponível para ser recuperada por ferramentas baseadas em metadados, marcas e elos estabelecidos no conteúdo da notícia. Os objetivos deste artigo são apresentar e discutir alguns problemas e soluções relacionadas à autoria de notícias hipermídia/multimídia. Aspectos como produção de conteúdo jornalístico multimídia (Jornalismo Multimídia), produção de notícias para apresentação em diversos meios de comunicação, preservação da semântica do conteúdo das notícias e o estabelecimento de ligações entre notícias correlatas são discutidos no artigo. Um modelo para representar e gerenciar notícias, baseado em linguagens de marcação, é apresentado junto com uma ferramenta projetada para minimizar possíveis resistências que os jornalistas possam oferecer a estas mudanças.*

1. Introdução

Os modelos atuais de produção de notícias têm sido direcionados preferencialmente para a publicação. Devido a este enfoque, um importante acervo jornalístico utilizado ou não para a produção das matérias publicadas é simplesmente perdido. Por falta de espaço na edição corrente, diariamente milhares de notícias, oriundas das mais diversas fontes, tais como agências nacionais, internacionais, colaboradores e jornalistas locais, são descartadas [Prazeres 2002]. As notícias publicadas, e nesse caso arquivadas e indexadas (conhecidas por “*old news*”), também não são encaradas como fonte alternativa de receita pela maioria das empresas do ramo. Por este motivo, os mecanismos desenvolvidos para a representação, armazenamento e exploração desse material por jornalistas, historiadores, pesquisadores e por novas aplicações computacionais, estão longe de serem satisfatórios [Schaffer 2000].

O domínio, por parte dos profissionais geradores de notícias, dos processos de produção de conteúdos (independentemente da mídia utilizada) e de composição desses diversos conteúdos e mídias para a apresentação de uma determinada notícia possibilitará que os avanços tecnológicos nas áreas da multimídia façam parte do cotidiano desses profissionais. O processo de produção de conteúdo jornalístico deverá ser adequado a este novo tipo de jornalismo, contemplando as diversas mídias existentes (como vídeo, áudio, gráfico, imagem, etc.) e as diversas formas de expressar uma mesma notícia utilizando diferentes meios de apresentação. É exatamente este tipo de produto jornalístico, resultante deste novo processo produtivo, que deve possibilitar a chamada convergência das mídias para a comunicação.

Com a convergência das mídias, com o desenvolvimento da tecnologia da informação e, sobretudo, com a utilização da Internet como meio de comunicação, surge a possibilidade e a necessidade do desenvolvimento do Jornalismo Multimídia.

Jornalismo Multimídia é entendido aqui como a produção de conteúdo jornalístico para diversos tipos de objetos mídia (texto, áudio, imagem, vídeo), possibilitando a sua distribuição em diversos meios (ou veículos) de comunicação, tais como jornal, revista, rádio, TV e a Web. A produção de conteúdo leva em conta não só as peculiaridades da forma de apresentação de cada um desses meios de comunicação, mas também a linguagem adequada a cada um deles.

A ação da tecnologia e, sobretudo a Web, têm proporcionado uma convergência dos processos de produção e difusão das informações. Entretanto, esta convergência esbarra num componente essencial do sistema: o jornalista, elemento responsável pela produção do conteúdo. Jornalistas treinados e adaptados a um tipo de meio de comunicação (jornal, por exemplo) não têm, na maioria das vezes, a experiência necessária para trabalhar com outros meios (rádio ou TV), seja pela diferença na forma de produção dos conteúdos, seja pela linguagem a ser utilizada. Por motivos análogos, a utilização de diversos objetos mídia na produção de uma notícia é ainda uma tarefa muito complexa.

A padronização da representação de conteúdos jornalísticos através de linguagens de marcação baseadas em XML tem como consequência a melhoria no processo de gerência do ciclo de vida de notícias. Linguagens de marcação, como as utilizadas neste artigo, podem também dar suporte ao Jornalismo Multimídia por serem linguagens que independem tanto da mídia utilizada como do meio de apresentação.

São muitos os benefícios da adoção de linguagens padrões para representação de notícias. Grande parte do trabalho de formatação e conversão de um conteúdo para adaptá-lo ao meio de apresentação (TV, Web, rádio, jornal escrito, etc.) pode ser eliminada se o conteúdo original da matéria for gerado com versões específicas de apresentação para o meio a qual se destina. Mais ainda, esta conversão pode ser automatizada de forma a entregar o documento, recém-preparado ou recebido de uma fonte qualquer, para a edição específica para o meio que fará efetivamente a sua distribuição.

Além disso, a mudança no processo de produção de notícias para diferentes meios de apresentação possibilita o surgimento de novos e interessantes serviços para o usuário final da notícia (o leitor), bem permite que se preserve a qualidade e a semântica do produto final (a notícia). Um serviço que pode ser implementado de forma efetiva a partir deste novo processo de produção de notícias é o chamado “*Daily Me*”. Neste serviço, o usuário recebe apenas as notícias (ou informações, de maneira geral) que quer consumir (resultado do processamento de rigorosa especificação feita pelo usuário), na quantidade e qualidade que respeitam suas preferências e seus recursos de comunicação, processamento e apresentação [Allday 2001].

O surgimento do Jornalismo Multimídia aponta ainda para a necessidade de mudança no perfil dos profissionais da área. Neste contexto, os jornalistas aparecem como geradores de conteúdo no sentido mais amplo, deixando de lado as especificidades e peculiaridades de cada meio de comunicação. É possível que aconteça uma certa resistência a este novo processo por parte dos jornalistas, visto que toda mudança é sempre trabalhosa até que se adapte a ela. Com o objetivo de facilitar esta mudança de paradigma da geração de conteúdos jornalísticos, o artigo propõe uma metodologia para produção de notícias multimídia voltadas para diversos meios de apresentação e uma ferramenta de suporte a esta metodologia. O objetivo desta metodologia é permitir a implementação destas mudanças dentro de um ambiente de geração de notícias, sem que para isso o dinamismo e a criatividade, inerentes ao trabalho jornalístico, sejam destruídos.

Foge ao escopo deste artigo a discussão sobre a capacitação do novo perfil profissional de jornalismo. Pode-se, porém sugerir a criação de equipes compostas por profissionais com as diversas habilidades, assim como já se faz na produção de infográficos, até que no futuro torne-se comum a formação de profissionais multimídia.

O artigo está organizado em 7 seções. A próxima seção faz uma breve descrição das linguagens de marcação para a representação, gerenciamento e intercâmbio de notícias multimídia/hipermídia. A seção 3 apresenta a metodologia que propomos para a produção de notícias. A seção 4 descreve uma ferramenta de suporte à metodologia proposta. A seção 5 apresenta algumas considerações e conclusões importantes do trabalho desenvolvido. E por fim agradecimentos e referências bibliográficas.

2. Linguagens de Marcação para Notícias

Diversos padrões no domínio das notícias, baseados em XML e recentemente propostos, podem ser destacados, tais como: PRISM [PRISM], XMLNews [XMLNews], NewsML [NewsML] e NITF [NITF].

PRISM (*Publishing Requirements for Industry Standard Metadata*) é um vocabulário de metadados para dar suporte ao processo de produção de publicações e intercâmbio de conteúdos para revistas, catálogos, livros e jornais. Funciona como um padrão para descrição de conteúdo, intercâmbio e reutilização tanto de publicações tradicionais (papel) quanto em publicações eletrônicas. PRISM provê um arcabouço (*framework*) para intercâmbio e preservação de conteúdo e também, metadados que definem uma coleção de elementos para descrever este conteúdo e um conjunto de vocabulários controlados listando os valores para estes elementos. É um padrão utilizado e desenvolvido pela IDEAlliance (*International Digital Enterprise Alliance*) [IDEAlliance].

Proposta em 1998, a linguagem XMLNews foi uma tentativa de se estabelecer um padrão para descrição de notícias em formato XML. Ela é dividida em duas partes: (i) XMLNews-Story, baseada numa versão do NITF de 1998, utilizada para atribuir marcações no texto (conteúdo) da notícia; e (ii) XMLNews-Meta, utilizada para descrever o conteúdo (título, linguagem, direitos, revisões, publicação, assunto, relacionamento entre as notícias, etc.). Esse padrão acabou caindo em desuso devido à evolução do NITF e o surgimento do NewsML.

NewsML é uma linguagem e um modelo estrutural, baseado em XML e independente de mídia, adequado para a representação de notícias nos vários estágios de seu ciclo de vida em um ambiente de serviço eletrônico. Pode ser utilizado tanto em sistemas editoriais quanto para o intercâmbio entre provedores de serviço de notícias e consumidores finais. Foi criada e desenvolvida pelo IPTC [IPTC] (*International Press Telecommunications Council*) e atualmente é utilizada e/ou apoiada pelas principais agências de notícia do mundo, como Reuters [Reuters], AFP [AFP], NAA [NAA].

NITF é uma linguagem específica para descrever conteúdo jornalístico no formato texto. Também criado e desenvolvido pelo IPTC, tornou-se o padrão para o texto em notícias no formato NewsML.

NewsML e NITF são duas linguagens que foram projetadas especificamente com o propósito de enriquecer e facilitar a manutenção de parte da semântica da enorme quantidade de informações produzidas e armazenadas pelas agências de notícias. São hoje as linguagens mais importantes, mais completas e de melhor aceitação na indústria de notícias. Além de viabilizar o gerenciamento de todo o ciclo de vida de uma notícia e de facilitarem o intercâmbio entre agências de notícias, NewsML e NITF permitem a atribuição de metadados às notícias, o relacionamento entre as notícias, etc. A metodologia e a sua ferramenta de apoio propostas no artigo estão fundamentadas nestas duas linguagens de marcação para notícias, apresentadas com mais detalhes nas seções 2.1 e 2.2.

2.1. NewsML

A linguagem NewsML foi concebida para uso na produção, no armazenamento e na distribuição eletrônica de notícias, não se atendo, portanto, às necessidades específicas da publicação em papel (muito embora possa dar todo o suporte ao jornal em papel durante as fases em que a notícia é manipulada eletronicamente).

Do ponto de vista de modelo estrutural, NewsML suporta diferentes objetos mídia e formatos, além de permitir a estruturação hierárquica de uma notícia em torno

de itens de notícia, de seus relacionamentos e seus metadados. Apesar de ser independente de mídia, NewsML fornece mecanismos específicos para manipulação de texto, para o qual é utilizado o padrão NITF.

No padrão NewsML, o conteúdo de uma matéria jornalística é estruturado sob a forma de um documento XML, independentemente do meio em que eventualmente a notícia venha a ser veiculada. O documento empacota todas as informações disponíveis relacionadas ao tema da notícia. Estas informações podem ser incluídas sob diferentes tipos de modalidades, formatos, linguagens e codificações, facilitando o intercâmbio entre agências de notícia, empresas de comunicação e trazendo vantagens para sua manipulação.

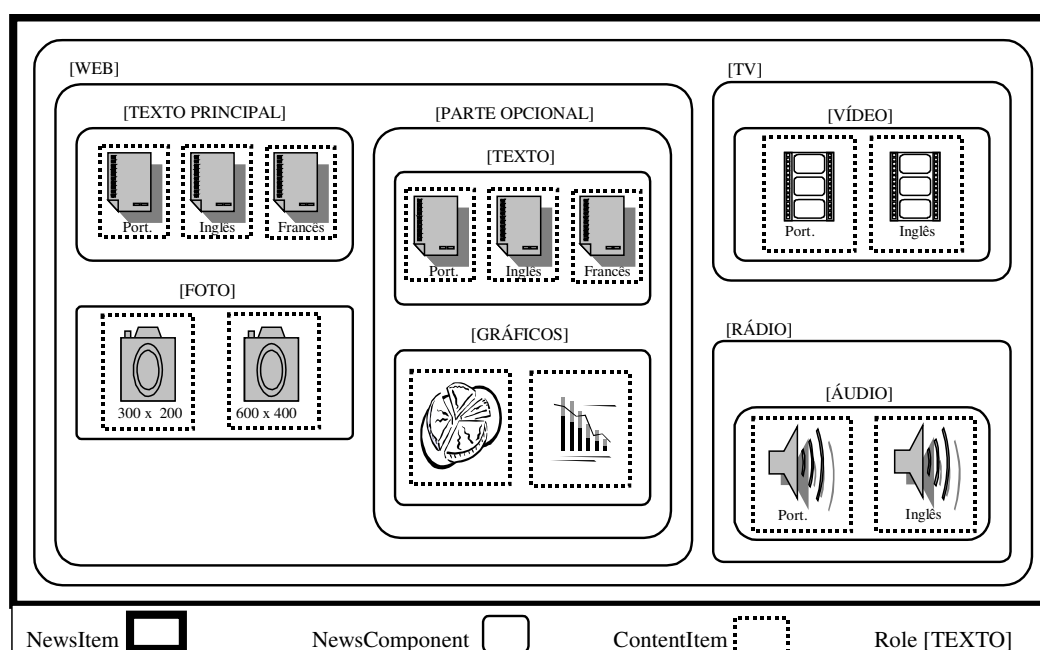


Figura 1. Conteúdo empacotado em NewsML.

Um documento NewsML funciona como um pacote contendo uma notícia, seus metadados, mídias e suas características. A visão de conteúdo empacotado viabiliza a estruturação de um item de notícia visando o seu armazenamento e intercâmbio entre organizações que utilizam o NewsML como padrão. A Figura 1 ilustra uma notícia empacotada, composta de diversos objetos (geralmente referências — URLs onde encontram-se os dados correspondentes) e organizada para apresentação em três veículos diferentes: Web, TV e rádio.

NewsML utiliza vocabulários controlados para definir o valor de determinados elementos (marcações), ou seja, esses elementos terão um “domínio” definido por um vocabulário controlado. Isso torna NewsML bastante flexível, pois cada agência de notícias ou jornal pode ter seu próprio vocabulário controlado, ao mesmo tempo impor um limite à escolha dos valores, evitando que sejam criados documentos com diversos termos diferentes para expressar o mesmo significado (semântica). Pode-se ter, por exemplo, um vocabulário controlado para definir o *status* da notícia; os valores possíveis para o elemento *Status* de NewsML definidos por um determinado vocabulário controlado poderiam ser: publicável, embargada, retida e cancelada.

Os metadados de NewsML permitem armazenar as diversas propriedades de objetos de notícia. Existem três categorias de metadados em NewsML, que são definidas de acordo com suas funções específicas, utilizando os seguintes elementos:

- `AdministrativeMetadata` – guarda informações sobre o editor, criador, origem, colaboradores e histórico de um objeto de notícia.
- `RightsMetadata` – guarda informações sobre o proprietário e direitos de uso de um objeto de notícia.
- `DescriptiveMetadata` – mantém informações sobre a linguagem, gênero, assunto, contexto, menção, com o que está relacionado e quem tem interesse por determinado objeto de notícia.

2.2. NITF

Um item de notícia do modelo NewsML, conforme visto na seção 2.1, pode ser composto de várias partes e por mídias distintas, ou seja, um item de notícia é um componente multimídia. Para a representação dos componentes da mídia texto, o vocabulário e o modelo NITF do IPTC podem ser utilizados em NewsML.

Tabela 1 – Principais elementos de semântica da linguagem NITF.

| Elemento | Funcionalidade |
|--------------------------|--|
| <code>person</code> | usado para destacar uma pessoa em um artigo |
| <code>name.given</code> | destaca o primeiro nome de uma pessoa |
| <code>name.family</code> | identifica o sobrenome de uma pessoa |
| <code>org</code> | usado para destacar uma organização sendo ela privada ou pública |
| <code>location</code> | local importante mencionado no artigo |
| <code>region</code> | destaca uma área geográfica |
| <code>money</code> | usado para identificar unidades monetárias |
| <code>num</code> | para destacar dados numéricos |
| <code>chron</code> | identifica referências à data e ao tempo |
| <code>event</code> | evento é uma atividade considerada interessante para ser publicada |
| <code>evloc</code> | local de ocorrência de um evento |
| <code>function</code> | cargo ou função exercido por uma pessoa |
| <code>lang</code> | usado para separar mudanças de linguagem em um texto |

Os elementos e atributos definidos por NITF permitem realizar uma marcação detalhada do conteúdo, capaz de viabilizar aplicações que possam explorar o significado da notícia. O conjunto de marcas das principais formas de preservação de semântica a um conteúdo NITF é apresentado na Tabela 1. Existem marcas definidas para a identificação de pessoas, locais, datas, nomes de empresas e organizações, valores, títulos, funções, eventos, etc. Deste modo, torna-se possível estabelecer uma marcação capaz de viabilizar a identificação de quem escreveu a matéria, quem pode revisá-la, sobre quem se fala na matéria, qual o assunto tratado, que evento é coberto, sobre que organização, quando e onde o fato ocorreu, entre outras informações.

3. Produzindo Notícias Estruturadas

Com o intuito de facilitar a ação do jornalista na criação de notícias multimídia, este artigo apresenta uma metodologia¹ que envolve a representação e o gerenciamento de notícias, usando os padrões NewsML e NITF. Esta metodologia estabelece uma série de passos que vão desde a geração de uma nova notícia, passando por suas eventuais modificações no decorrer de seu ciclo de vida, até a definição das formas de apresentação ao usuário final, procurando utilizar uma linguagem bastante clara e intuitiva para o jornalista.

A metodologia é apresentada esquematicamente na Figura 2. Ela consiste na seguinte seqüência de procedimentos a serem executados na criação/edição de uma notícia:

1. **Identificação do jornalista:** procedimento simples e efetuado com o uso de um identificador único de usuário e senha;
2. **Edição de notícias:** uma vez identificado, o jornalista pode executar dois procedimentos, o de criação de uma notícia ou a alteração (modificação) de uma notícia já existente:
 - a. **Criação de uma notícia:** o ponto de partida para este procedimento é a inserção de um conjunto de elementos NewsML e seus respectivos valores definidos por um vocabulário controlado e sugeridos a partir das informações associadas ao perfil do usuário (jornalista esportivo, policial, econômico, etc). Estes pares elementos-valores² (por exemplo, `<subject>economy</subject>`) são acrescentados ao conteúdo da nova notícia e podem ser ajustados caso os valores sugeridos não estejam exatamente de acordo com a notícia a ser criada. O jornalista pode fazer as devidas correções e ajustes nos metadados (sempre utilizando os valores pertencentes ao vocabulário controlado) de forma a adaptá-los à notícia trabalhada (por exemplo, alterar o valor de `<subject>` de *economy* para *sports*);
 - b. **Alteração de uma notícia:** nesse procedimento, os valores de elementos NewsML já associados à notícia a ser alterada são carregados para o ambiente de edição. O jornalista, partindo do conteúdo marcado, poderá fazer as modificações desejadas na notícia alterando os valores dos elementos de acordo com o vocabulário controlado. Ainda nesta fase, um documento de revisão é gerado para armazenar as modificações realizadas na notícia. As alterações são feitas sobre o documento original e todas as modificações efetuadas são armazenadas no documento de revisão. O documento de revisão funciona como um arquivo de histórico (*log*), o qual é marcado através de elementos NewsML específicos para revisão. Note que este arquivo de histórico sempre estará vinculado às próximas alterações (revisões) da notícia;

¹ Uma versão preliminar desta metodologia foi apresentada [Prazeres 2002].

² Estes pares elementos-valores definem os metadados padrões (iniciais) de uma notícia em NewsML.

3. **Inserção/remoção de componentes:** finalizado a etapa da associação dos pares elementos-valores, o jornalista deve passar à fase de inserção/remoção de componentes (objetos mídia) à notícia. A inserção/remoção destes componentes leva em conta o meio de apresentação ao qual a notícia se destina. Os componentes da notícia são marcados e estruturados conforme regras definidas em NewsML e conceito de conteúdo empacotado (garantindo a independência entre o conteúdo e a forma de apresentação), conforme ilustrado na Figura 1;

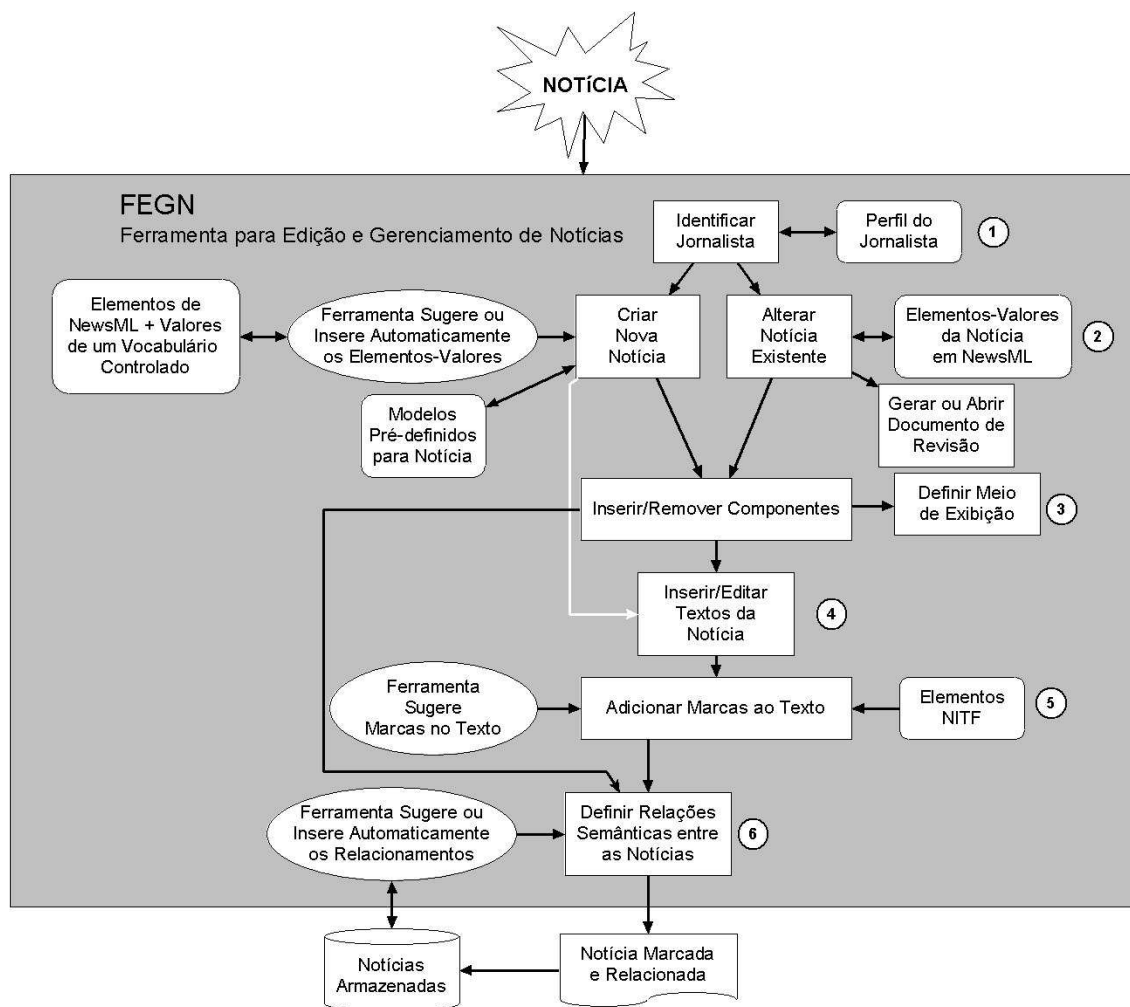


Figura 2. Visão Geral da Metodologia.

4. **Manipulação de componentes do tipo texto:** uma vez que a maior parte da semântica da notícia está associada a objetos mídia do tipo texto, a manipulação destes componentes é feita de forma um pouco diferente das outras mídias. Além disso, a marcação de conteúdos textuais de uma notícia geralmente utiliza o padrão NITF, definido especificamente para estes componentes. Assim as mídias do tipo texto são inseridas em separado das demais. Nesta fase o jornalista está restrito a utilizar o conjunto de marcas (elementos) NewsML e NITF.
5. **Inserção de semântica no texto:** novas marcações do conteúdo, em conformidade com o NITF, podem ser efetuadas pelo jornalista. O objetivo destas novas marcações no texto é a preservação da semântica do seu conteúdo

utilizando, para isso, os elementos listados na Tabela 1. Algumas marcações podem ser sugeridas ao jornalista que deve avaliar a possibilidade de efetuar-las;

6. **Definição de relações semânticas:** na fase final, o jornalista deve definir as relações semânticas entre a notícia que está criando/editando com outras notícias já existentes. Novamente o jornalista recebe sugestões sobre notícias que possam estar relacionadas com a notícia que ele está criando ou editando.

De acordo com esta metodologia a marcação do conteúdo pelo jornalista durante a criação/edição de uma notícia torna-se transparente. A Tabela 2 descreve os principais aspectos abordados na metodologia com o objetivo de facilitar a geração de conteúdos marcados pelo jornalista.

Tabela 2. Aspectos da metodologia que facilitam a ação do jornalista.

| Característica | Descrição |
|------------------------------------|--|
| Perfil do jornalista | Possibilita associar grande parte dos metadados automaticamente à notícia a partir de um perfil pré-definido do jornalista. |
| Modelos pré-definidos | Uma notícia pode ser criada a partir de um modelo pré-definido. Assim, o jornalista não precisa executar a tarefa de montar a estrutura da notícia. |
| Sugestão de marcas no texto | São sugeridas marcações de acordo com algumas características do texto, tais como: datas, números, unidade monetária, nomes (começados por letra maiúscula), etc. |
| Sugestão de relacionamentos | São sugeridas outras notícias, a partir dos metadados, que podem estar relacionadas com a notícia que está sendo criada. Alguns dos relacionamentos são criados automaticamente sem a intervenção do jornalista. |

4. FEGN: Ferramenta para Edição e Gerenciamento de Notícias

A partir da metodologia proposta na seção anterior surgiu a idéia da criação de uma ferramenta que apoiasse este novo processo de produção de notícias multimídia e que, ao mesmo tempo, facilitasse o trabalho do jornalista. A ferramenta FEGN suporta a edição e o gerenciamento de notícias marcadas de acordo com os vocabulários NewsML e NITF e de maneira semi-automática. Devido ao fato de que as principais API's (*Application Program Interface*) para manipulação de documentos XML e NewsML estarem disponíveis em Java, tornou-se natural a escolha desta linguagem para a implementação. As API's utilizadas foram: DOM [DOM], SAX [SAX], JDOM [JDOM] e NewsML Toolkit [NewsML Toolkit].

Ao utilizar a ferramenta, o jornalista não percebe que o produto final de seu trabalho é uma notícia no formato XML (NewsML e NITF) uma vez que ele não insere nenhuma marcação XML diretamente na notícia. A marcação do conteúdo é feita a partir do preenchimento de campos de formulários (ver Figura 3) que, na maioria das vezes, são preenchidos com sugestões pela própria ferramenta, e em outras vezes sem intervenção alguma, tornando o processo transparente e simples para o jornalista. Ao mesmo tempo ele consegue perceber que a notícia gerada é uma notícia multimídia, pois, existe a possibilidade de se adicionar diversos tipos de mídia à notícia. As seguintes funcionalidades estão implementadas na ferramenta:

- a. **Criação/Edição do perfil do jornalista:** um perfil é associado ao jornalista, contendo informações como nome, gênero (esportes, economia, política, etc.), assunto, etc.

- b. **Criação da estrutura da notícia:** o jornalista pode montar a estrutura da notícia vinculada ao meio de apresentação conforme ilustrado na Figura 1 e na Figura 3 que mostra a estrutura de uma notícia diretamente da ferramenta FEGN;
- c. **Um *wizard* com modelos para a estrutura da notícia pré-configurados em NewsML:** a partir dos modelos, o *wizard* guiará o jornalista na criação da notícia. Com o uso de modelos todo o trabalho de estruturação (item b acima) da notícia é eliminado. Os modelos foram definidos em função do meio de comunicação onde a notícia será veiculada. Dentre as opções de modelos que podem ser implementados estão um modelo para Web, um para jornal escrito, um para rádio, um para TV. Além destes, podem ser criados modelos combinando dois meios, tais como Web e jornal escrito, ou ainda TV e rádio. Finalmente, existe a opção do jornalista criar um modelo personalizado para a edição da notícia;
- d. **Associação automática de metadados à notícia sem intervenção do jornalista:** data de criação, informações sobre revisão, provedor da notícia, autor (a partir dos dados do jornalista que usa a ferramenta), etc;
- e. **Associação semi-automática de metadados à notícia com intervenção mínima do jornalista:** neste caso, os metadados são sugeridos pela ferramenta baseando-se no perfil do jornalista que está criando/editando a notícia. Este perfil define um domínio típico de atuação do jornalista e os possíveis elementos e relações dentro do mesmo;
- f. **Associação de semântica ao texto da notícia:** aqui, a ferramenta sugere ao jornalista possíveis marcações a serem efetuadas no documento utilizando o conjunto de marcas NITF, cujas marcas principais estão listadas na Tabela 1. Algumas marcações poderiam ser sugeridas, por exemplo, para as palavras do texto que comecem por letra maiúscula e não são início de parágrafo ou frase. Estas palavras estão normalmente associadas ao nome de uma pessoa, cidade, estado, país ou organização. Outra possível sugestão de marcação estaria relacionada às datas, unidades monetárias ou unidades de medidas encontradas no texto da notícia;
- g. **Criação automática de relacionamentos com notícias existentes:** a partir de metadados que sugerem relações evidentes entre as notícias como por exemplo:
 - i. nome do autor – relacionar notícias de um mesmo autor.
 - ii. data de publicação – relacionar notícias publicadas no mesmo dia.
 - iii. gênero – relacionar notícias do mesmo gênero.
 - iv. assunto – relacionar notícias do mesmo assunto.
- h. **Sugestões de outros possíveis relacionamentos:** o jornalista decide se as notícias estão relacionadas ou não. Um exemplo prático seria o caso da continuidade de uma notícia, onde a ferramenta sugere um conjunto de notícias do mesmo gênero e/ou assunto publicadas anteriormente (um período de três dias por exemplo) e a partir deste conjunto o jornalista escolhe as notícias que realmente estão relacionadas com a notícia atual.

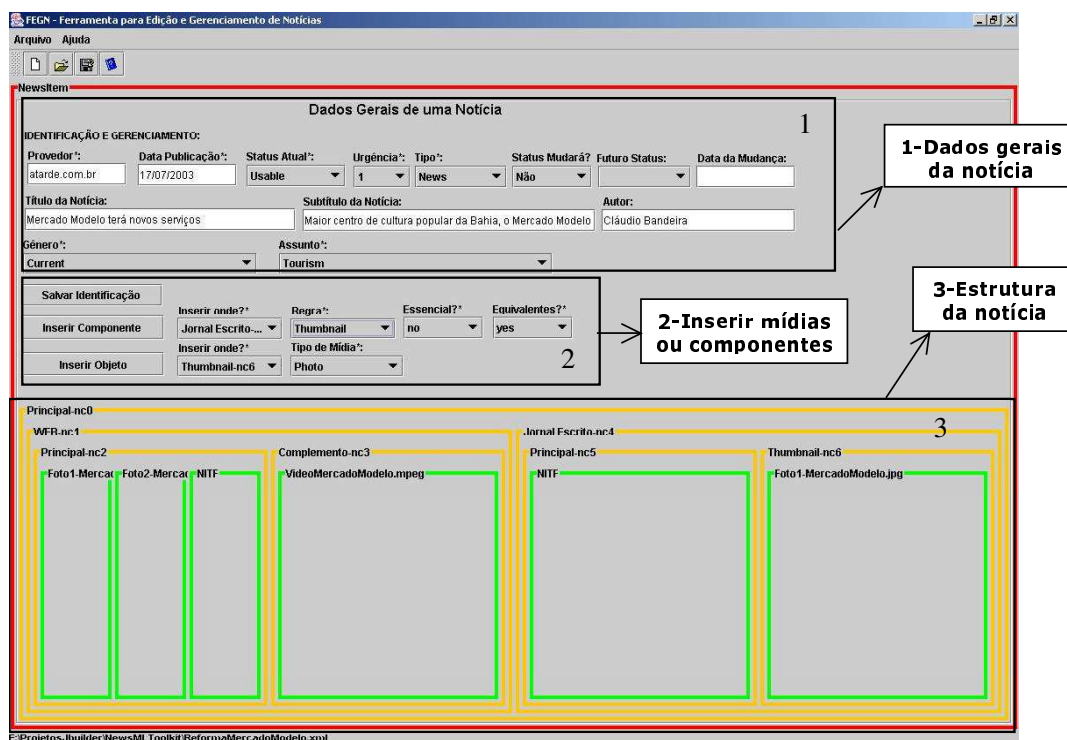


Figura 3. Estrutura de uma notícia visualizada na ferramenta FEGN.

A interface da ferramenta FEGN foi desenvolvida baseada no conceito de conteúdo empacotado de NewsML (Figura 1). Desta forma é possível que o jornalista veja a estrutura da notícia na medida em que os objetos vão sendo inseridos. A FEGN possui um formulário principal, que podemos dividir em três partes conforme ilustrado na Figura 3:

- 1- **Dados gerais da notícia:** trata-se de um pequeno formulário que deve ser preenchido com alguns dados iniciais a respeito da notícia, tais como título, subtítulo, autor, status, data de publicação, gênero, assunto, etc. Estes dados, na sua maioria, são preenchidos automaticamente ou sugeridos pela ferramenta a partir do perfil do jornalista conforme descrito anteriormente nesta seção.
- 2- **Inserir mídias ou componentes:** aqui se encontram os controles para inserção de componentes e objetos de notícias (mídias). A Figura 4 mostra o formulário de inserção/edição de um componente multimídia, neste caso uma imagem. Também neste formulário a ação do jornalista é mínima, pois, ele apenas deve informar o arquivo correspondente à mídia que será inserida.
- 3- **Estrutura da notícia:** uma visualização da estrutura da notícia, estrutura essa que foi criada pelo jornalista ou simplesmente gerada a partir de um modelo pré-definido. Em tempo de criação/edição da notícia o jornalista tem uma visão geral de toda a notícia conforme mostrado na Figura 3, onde a estrutura da notícia em NewsML é formada por um componente principal que contém dois outros componentes (Web e jornal escrito) que por sua vez contém outros componentes que mantêm os objetos da notícia (fotos, textos, vídeos, etc.) específicos para cada meio de apresentação.



Figura 4. Formulário para inserção/edição de mídias.

Além de criar/alterar notícias seguindo a metodologia que propomos na seção 3 deste artigo, a ferramenta FEGN possibilita importar notícias oriundas de outras agências de notícia que utilizam o padrão NewsML. Nos testes realizados foi possível importar e alterar notícias das agências Reuters e AFP, duas das maiores agências do mundo e que estão bastante engajadas no uso do NewsML.

5. Conclusões

O artigo apresenta uma metodologia para o desenvolvimento de hiperdocumentos baseada nas linguagens de marcação NewsML e NITF. A aplicação desta metodologia deve permitir a inclusão da tecnologia XML e seus benefícios dentro de um ambiente de criação e edição de notícias. Como resultado, estimula-se a prática de produção de notícias (conteúdo digital) conforme padrões internacionais facilitando a criação, edição, agregação de semântica ao conteúdo e a sua distribuição sob diferentes meios.

Uma ferramenta de apoio à geração de conteúdos marcados foi desenvolvida: a FEGN. Esta ferramenta é parte fundamental da metodologia, simplificando a tarefa de criação de conteúdos jornalísticos conforme os padrões NITF e NewsML. Buscou-se com esta ferramenta diminuir ao máximo o trabalho do jornalista na criação de uma notícia multimídia/hipermídia, de forma que as resistências a este novo processo de produção de notícias fossem as mínimas possíveis por parte dos jornalistas.

No momento em que este artigo foi escrito nenhum teste da ferramenta FEGN foi realizado em um ambiente típico de produção de notícias (como um jornal, por exemplo). Temos o apoio de um jornal de grande circulação no estado da Bahia (Jornal A Tarde) e os testes devem ser efetuados no ambiente de produção de notícias deste jornal. Para isto será de suma importância a colaboração de jornalistas e editores para que possamos testar a viabilidade da ferramenta bem como realizarmos as devidas correções e ajustes que se façam necessários.

Um trabalho semelhante está sendo desenvolvido por pesquisadores chineses da *Chinese NewsML Community* [CNewsML], uma comunidade chinesa que teve origem na Universidade de Hong Kong. O trabalho dessa comunidade baseia-se no desenvolvimento da ferramenta CIDAX (*Corporate Information Digital Archives in XML*). Esta ferramenta provê três principais funcionalidades: (i) composição de pacotes de informações baseadas em NewsML; (ii) armazenagem e intercâmbio de documentos e pacotes NewsML e; (iii) busca e transformação de informações NewsML. Assim,

enquanto a ferramenta FEGN proposta neste artigo tem como foco a produção de conteúdo jornalístico em NewsML, mais precisamente a produção de documentos NewsML, a ferramenta CIDAX tem como objetivo principal a organização, distribuição e busca deste conteúdo.

Outro aspecto relevante do projeto, não abordado neste texto, refere-se ao modelo de armazenamento, que deve viabilizar a construção de aplicações eficientes que tirem proveito da preservação da semântica das notícias (ou pelo menos parte dela). Este aspecto está sendo abordado por outro trabalho desenvolvido em conformidade com o estudo apresentado neste artigo.

Os resultados deste trabalho podem ser usados como base por empresas e entidades manipuladores de material jornalístico em sua transição para um novo paradigma de produção de notícias, em que repórteres, críticos, editores, revisores, etc. passam a assumir também a responsabilidade pela preservação do significado do conteúdo.

6. Agradecimentos

Os autores são gratos ao suporte financeiro concedido pela FAPESB (Fundação de Amparo à Pesquisa do Estado da Bahia) e ao Jornal A TARDE (Salvador – BA) para a realização deste trabalho.

7. Referências

AFP – Agence France Press. < <http://www.afp.com/>>.

Allday, T. “NewsML – enabling a standards-led revolution in news publishing?”, EBU Technical Review, june 2001, 8p. < http://www.ebu.ch/trev_287-allday.pdf>.

CNewsML – Chinese NewsML Community. <<http://cnewsml.org/english/index.shtml>>.

DOM – Document Object Model. < <http://www.w3c.org/DOM/> >.

Fernández, L. S. et al. “Application of XML to the Journalism Field”, Upgrade: The European Online Magazine for the IT Professional, Vol. III, Nº 4, Aug. 2002, pp. 38-42.

IDEAlliance – International Digital Enterprise Alliance. <<http://www.idealliance.org/>>.

IPTC – International Press Telecommunications Council. <<http://www.iptc.org/>>.

JDOM. <<http://www.jdom.org/>>.

Jokela, S. et al. “The Role of Structured Content in a Personalized News Service”, In: Proceedings of the 34th IEEE Hawaii International Conference on Sistem Science, 2001, p. 1-10.

NAA – Newspaper Association of America. < <http://www.naa.org/>>.

NewsML – News Markup Language. <<http://www.newsml.org/>>.

NewsML Toolkit. <<http://newsml-toolkit.sourceforge.net/>>.

NewsML Version 1.1-Functional Specification. International Press Telecommunications Council, Oct. 2002. < <http://www.iptc.org/site/NewsML/NewsMLSpec.htm> >

NITF – News Industry Text Format Documentation. <<http://www.nitf.org/>>.

NITF Tutorial. <<http://www.nitf.org/site/tutorial.html>>.

Prazeres, C. et al. “Uma Metodologia para a Produção e Suporte à Gerência de Hiperdocumentos Jornalísticos Fundamentada em Linguagens de Markup”, In: VIII Simpósio Brasileiro de Sistemas Multimídia e Hiperemídia. Anais do SBMIDIA, Fortaleza-CE, 2002, v. 1, p. 86-99.

PRISM – Publishing Requirements for Industry Standard Metadata. <<http://www.prismstandard.org/>>.

Reuters. <<http://www.reuters.com/>>.

SAX – Simple API for XML. <<http://www.saxproject.org/>>.

Schaffer, B. “Old news may be newspaper’s best reason to adopt XML”, Newspapers & Technology, 2000, pg. 36.

W3C – World Wide Web Consortium. <<http://www.w3.org/>>.

XMLnews – XML for News. <<http://www.xmlnews.org/>>.