

# WSPPD 2007

## *Um Modelo Para Conversão de Documentos de Texto Baseados em XML*

Caetano Sauer  
Nicolas B. Maillard

# Roteiro da apresentação

- Contextualização
- Formatos de arquivo baseados em XML
- Documentos de texto e o OpenXML
- Convertendo documentos para OpenXML

# Contextualização

- Trabalho desenvolvido no Núcleo de Desenvolvimento Open Source e Interoperabilidade do convênio UFRGS / Microsoft
- Estudo da norma Office OpenXML, proposta pela Microsoft e utilizada no Word 2007
- Prototipação de um conversor de documentos de texto para OpenXML, a partir de diversos formatos de entrada

# Formatos de arquivo baseados em XML

- Linguagem de marcação (*markup*): associa informações adicionais a um determinado texto

```
<informação> Texto </informação>
```

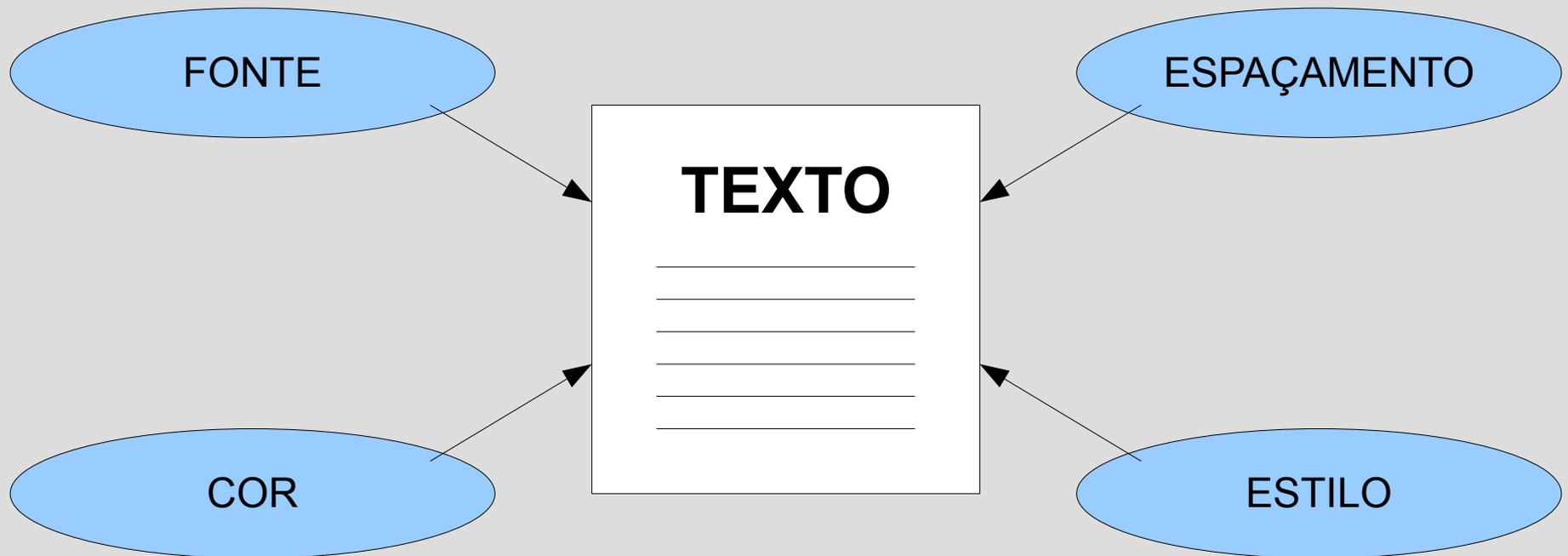
- XML serve de base para definir linguagens derivadas, em aplicações específicas, onde as informações carregadas pelo texto ganham significado
- Representa documentos através de uma estrutura hierárquica

# Formatos de arquivo baseados em XML

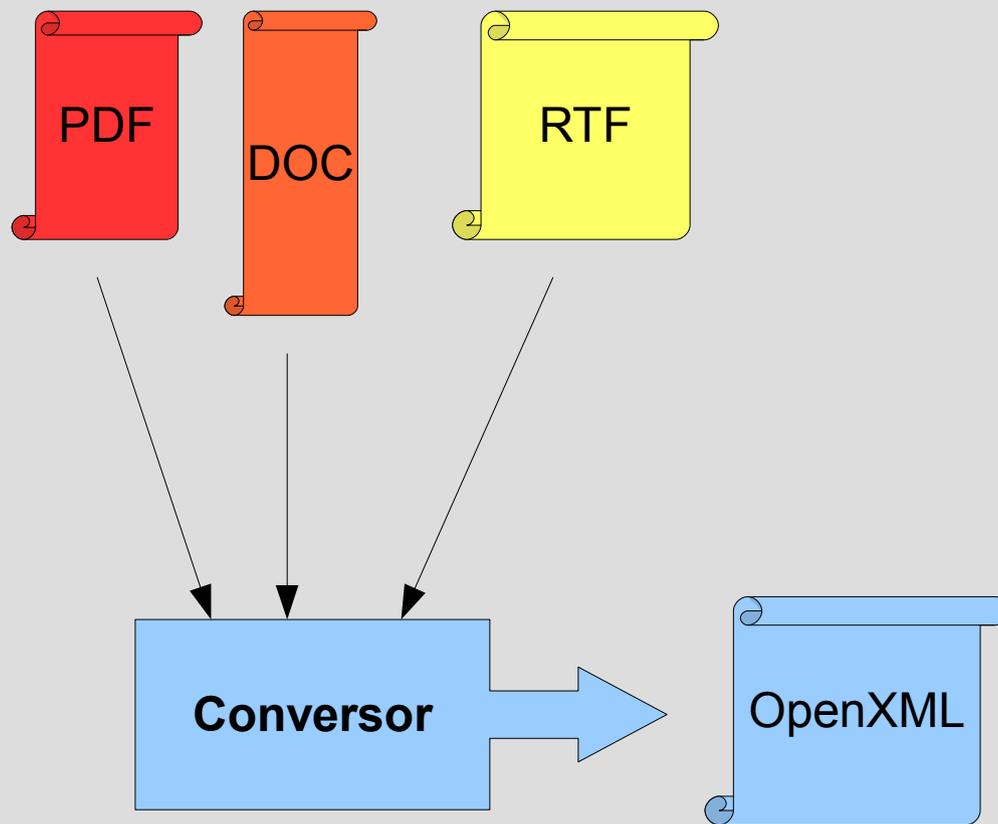
- Vantagens:
  - Legibilidade
  - Portabilidade
  - Suporte nativo a Unicode
  - Sintaxe universal
- Desvantagens:
  - Redundância de dados
  - Custos de armazenamento e processamento elevados
  - Estrutura hierárquica (em árvore) é uma limitação

# Documentos de texto e o OpenXML

- A noção de associar informações a um texto é natural em documentos de texto formatado

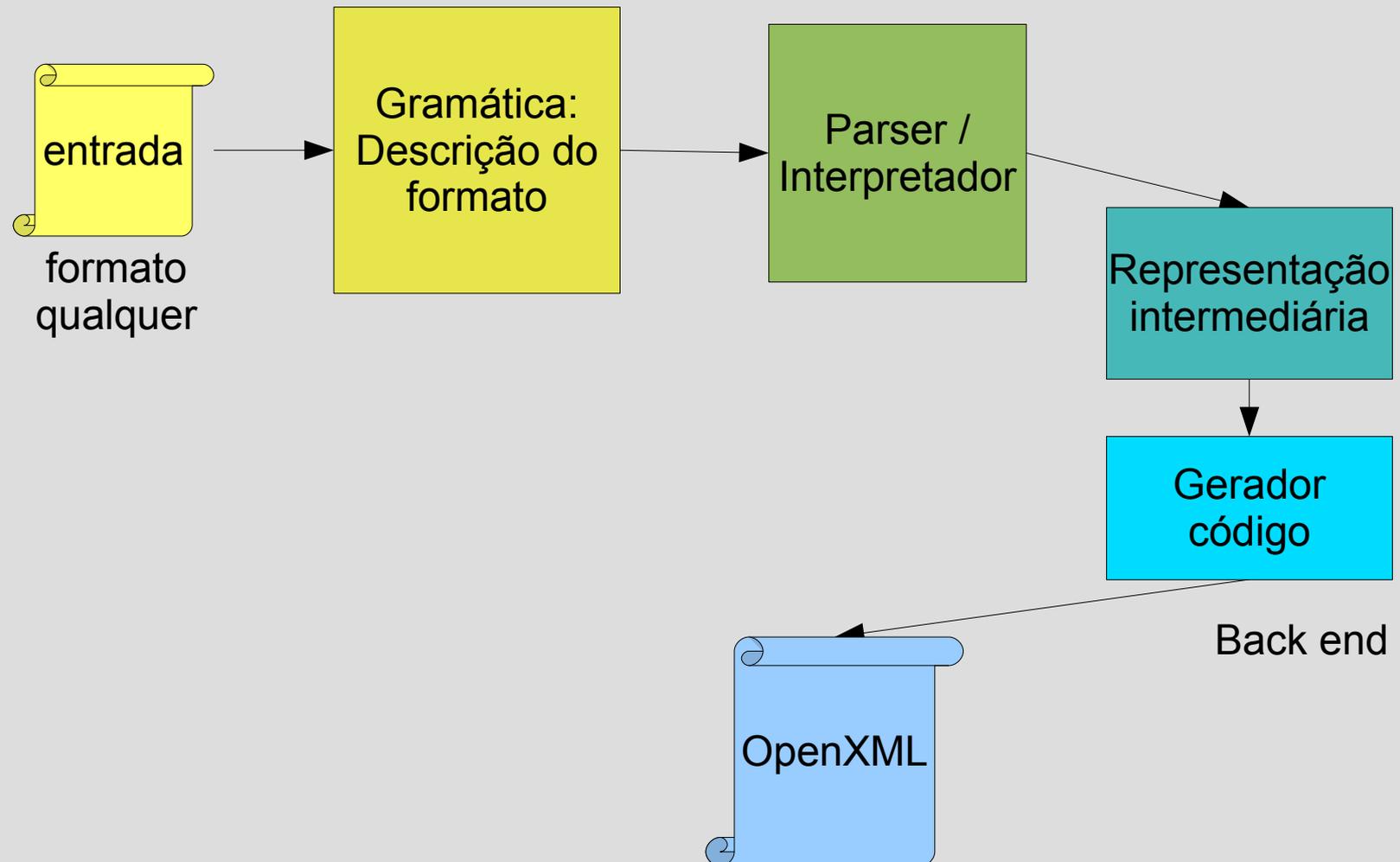


# Convertendo documentos para OpenXML

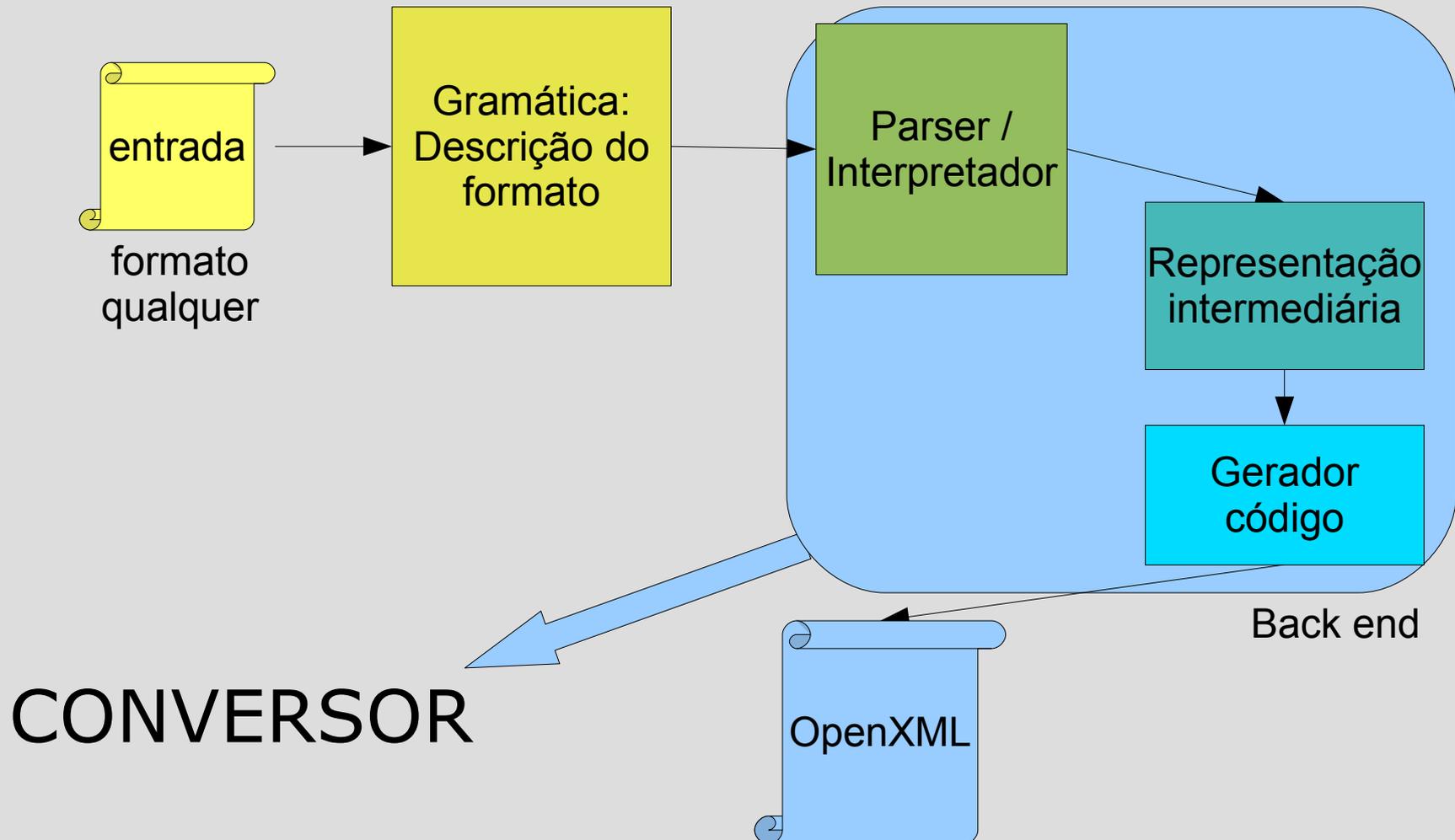


- Necessidades para construir o conversor:
  - Interpretadores para os formatos de entrada
  - Biblioteca para gerar os arquivos XML

# Convertendo documentos para OpenXML



# Convertendo documentos para OpenXML



# FIM

Página do projeto:

[www.codeplex.com/openxmlconv](http://www.codeplex.com/openxmlconv)