

Extração de informação de texto formal para a modelação expedita de edifícios monumentais

Ricardo Baptista
DEI - FEUP/ LIACC
Rua Dr. Roberto Frias, s/n
4200-465 Porto
ricardjose@fe.up.pt

Luís Paulo Reis
Universidade do Minho/LIACC
Guimarães/Porto
lpreis@dsi.uminho.pt

Luís Magalhães
Universidade de Trás-os-Montes e
Alto Douro/INESC TEC
Vila Real/Porto
lmagalha@utad.pt

Roberto Rodrigues António Coelho Augusto Sousa
INESC TEC / DEI - Faculdade de Engenharia, Universidade do Porto
Rua Dr. Roberto Frias, s/n, 4200-465 Porto
{roberto.rodrigues, acoelho}@fe.up.pt

Resumo

A extração de informação a partir de descrições textuais para a modelação procedimental de ambientes urbanos é apresentada com solução para os edifícios antigos. No entanto, este tipo de edifício carece de maior cuidado com os detalhes de alto nível. Este artigo descreve uma plataforma para a geração expedita de modelos 3D de edifícios monumentais, cuja arquitetura é modular. O primeiro módulo permite a extração de informação a partir de textos formais, pela integração do NooJ num Web Service. No segundo módulo, toda a informação extraída é mapeada para uma ontologia que define os objetos a contemplar na modelação procedimental, processo esse realizado pelo módulo final que gera os modelos 3D em CityGML, também como um Web Service.

A partir desta plataforma, desenvolveu-se um protótipo Web para o caso de estudo da modelação das igrejas da cidade do Porto. Os resultados obtidos deram indicações positivas sobre o modelo de dados definidos e a flexibilidade de representação de estruturas diversificadas, como portas, janelas e outras características de igrejas.

Palavras-chaves

Procedural modelling, natural language processing, information extraction.

1. INTRODUÇÃO

A modelação expedita apresenta-se como uma solução para a criação de ambientes urbanos com redução de custos e de tempo de execução. Contudo, para manter uma fidelidade visual aceitável é preciso aceder a fontes de informação, que no caso dos edifícios antigos, existe em grande quantidade em formato textual. No trabalho apresentado, damos o enfoque à extração do texto formal para um formato intermédio (XML) e, posterior conversão para o formato CityGML [Kolbe05]. Esta tecnologia é aplicada na modelação das igrejas do Porto.

2. TRABALHO RELACIONADO

Este trabalho segue uma linha de investigação relacionada com a extração de informação para edifícios monumentais [Rodrigues10]. A automatização do processo de modelação de edifícios baseia-se nas ferramentas automáticas de tradução e geração de resumos [Baeza04]. A utilização do NooJ neste trabalho revelou-se importante pelas seguintes características: multilingue, análise gramatical, sintática e morfológica de múltiplos textos em vários formatos, inserção de anotações ao texto e aplicação de gramáticas personalizadas [Silberstein05].

Para a visualização dos modelos 3D gerados, utilizou-se o CityGML, [Kolbe05].

3. ARQUITETURA DO SISTEMA

O sistema proposto designa-se por MEXEM (Modelação EXpedita de Edifícios – Monumentais) e é composto por três blocos: modelo de dados, extração e modelação3D.

3.1 Modelo de Dados

Entende-se por edifícios monumentais as catedrais, palácios e castelos cujas estruturas são complexas, precisando dum elevado nível de detalhe quando modelados.

A Figura 1 apresenta uma representação simplificada do modelo de dados desenvolvido, que faz corresponder os nós do ficheiro XML com os atributos das classes criadas para a criação do modelo em cityGML.

3.2 Extração de Informação com NooJ

A extração de informação tem como ponto de partida a extração de palavras-chave ou padrões de informação do ficheiro de texto. Para garantir a interoperacionalidade entre os vários módulos assume-se que o formato XML é o garante necessário, obedecendo às especificações do *schema* que suporta o modelo de dados.

