

Projectos Informáticos do Arquivo Distrital de Braga

MARIA DA ASSUNÇÃO JÁCOME DE VASCONCELOS
CLARA SOFIA DA ROCHA PINTO MOREIRA

Introdução

Com o presente artigo pretendemos descrever os objectivos e motivação do HITEX, um sistema informático desenvolvido no Arquivo Distrital de Braga (A.D.B./U.M.), a partir de 1988, em colaboração com o Departamento de Informática da Universidade do Minho e com o apoio da companhia IBM-Portuguesa e da Fundação Calouste Gulbenkian, para registar, organizar e gerir informação de natureza histórica.

Pretende-se, igualmente, dar a conhecer a colaboração do Arquivo Distrital de Braga em outros projectos informáticos com destaque para o HINET e, actualmente, no Geira.

HITEX: um sistema em desenvolvimento

O *software* a que se designou de HITEX teve como objectivo a melhoria do acesso dos utentes do Arquivo Distrital de Braga aos fundos documentais nele existentes, bem como a necessidade do Arquivo em melhorar a sua tecnologia de publicação de fontes.

Assim, a filosofia de base do HITEX foi criar um sistema com «uma arquitectura aberta capaz de trocar informação com ferramentas em uso» no Arquivo (processadores de texto e bases de dados já existentes) e em outros ambientes futuros, bem como a criação de um módulo para compilação automática de índices (onómástico, toponómico, ideográfico, sistemático, etc.), por se tratar de uma tarefa trabalhosa na edição crítica de fontes históricas e arquivísticas.

A primeira fase do projecto concentrou-se na utilização de uma linguagem fácil para transcrição documental cuja descrição poderá ser encontrada em «HITEX. Um sistema para transcrição documental em larga escala» [PEREIRA, 1991]. De salientar que, na preparação do texto optou-se como ferramenta de base do sistema HITEX, o sistema LATEX, pelas seguintes razões:

- o LATEX baseia-se numa linguagem gráfica estruturada, de grande capacidade expressiva;
- o LATEX está disponível em praticamente todos os recursos computacionais, desde IBM-PC compatíveis aos ambientes Macintosh e Unix;
- o LATEX produz resultados gráficos (em TEX) de muito boa qualidade.

Escolhido o texto base passou-se à segunda fase do projecto, ou seja, à tarefa da *representação do conhecimento*, que neste sistema resultará nos índices, sendo estes uma representação (textual) de conhecimento histórico e/ou arquivístico. Os índices são, pois, produtos de um sistema de representação do conhecimento em que a sua geração automática será feita a partir de relatórios textuais de uma base de conhecimento previamente construída. Quer isto dizer, que a arquitectura da base de conhecimento do HITEX assemelha-se a uma rede semântica, com classes e subclasses, ou seja, a uma taxonomia de classes. Na sua feitura, a classe é caracterizada por atributos (propriedades), que serão dados a arquivar sobre os componentes dessa classe, também designados por instâncias. Nesta estrutura interligada, ou em rede, que é a base do HITEX, reside uma grande economia e flexibilidade do módulo, em virtude de todo o tipo de informação ficar automaticamente disponível, podendo, se for caso disso, ser acrescentado pelo seu autor/produtor.

Do ponto de vista do utilizador, a interface do protótipo permite basicamente realizar as seguintes tarefas:

- transcrição de fontes documentais;
- acesso ao *facsimile* do documento;
- marcação do texto com a identificação de entradas no índice e descrição das ditas marcas.

As marcas a que nos referimos correspondem aos elementos (*instâncias*) da base de conhecimento e a sua descrição corresponde ao preenchimento das suas características (*atributos*), pelo que

todas estas tarefas visam a construção e enriquecimento da base de conhecimento.

Posteriormente, o utilizador poderá gerir esta base de conhecimento através de actividades como:

- navegação sobre o diagrama da taxonomia do sistema;
- procura de componentes;
- criação de um novo componente da classe correntemente seleccionada na taxonomia;
- preenchimento de novos dados referentes ao componente correntemente seleccionado;
- especialização de um dado componente;
- adição de uma nova classe à taxonomia;
- adição de mais atributos a uma dada classe;
- interpelação assistida da base de conhecimentos; etc.

As duas figuras seguintes mostram ecrãs do protótipo disponível em MS-DOS. A janela superior esquerda mostra a taxonomia do sistema. Nas janelas laterais faz-se o acesso e a gestão de atributos e instâncias. Toda a conversação com o utilizador é feita através de menus. A janela inferior tem a ver com transcrição de fontes. Esta janela não é mais do que um simples editor de texto, capaz de comunicar com a base de conhecimentos sempre que se quiser. É de notar que uma fonte pode ser transcrita usando outros editores de texto, sendo posteriormente importada para esta janela com facilidade desde que não contenha caracteres de controlo obscuros. Este editor tem a vantagem de permitir referências cruzadas em modo semi-automático, ou seja, deixando o cursor na posição pretendida o utilizador pode pesquisar a base procurando os componentes que lhe interessarem. Um simples toque no rato será suficiente para que os identificadores dos componentes seleccionados sejam inseridos automaticamente na janela de texto, imediatamente após a posição do cursor. *Relembra-se que tais identificadores têm por objectivo marcar a ocorrência de itens que se destinam a índices.*

HiTeK System - Semantic Net User Interface on :

Universo

- Cronologico
- Documental
- Ideografico
- Onomastico
 - Antroponimico
 - Toponimico

Attributes
 nome
 titulo

Types

Instances 81
 Antro109
 Antro111
 Antro112
 Antro113
 Antro114
 Antro313
 Antro314
 Antro315
 Antro341
 Antro37
 Antro38
 Antro39

Value of attribute: ?, in instance: Antro341

Instance Description: Antro341
 Antroponimico
 Cleópatra VII, Rainha do Egipto

Document : \stk\snet\festas
 \espacamento{1}
 \italico{\centrado{Se ao remo nos condenastes,}
 \centrado{Sendo t\~{a}o grande piloto,}
 \centrado{desculpado est\{a} \aAntro314 Fernando}
 \centrado{em dar com a barca nos lodos. }}
 \espacamento{1}

Current operation :

Figura 1 – Aspecto do ecrã no tratamento de uma fonte documental

Sistema HiTeK - Interface sobre a base de conhecimento: \stk\work\sfestas

Imagem : Index das Gavetas Vol.1 - F.1

Universo

Ua lo

Desc

Antro

Antó

Docu

nune

pagi

Extracção das Festas que os Cavalheiros de Braga fizeram aos Senhores Marquês de Aronches, Príncipe Senescal, Marquês de Vila Verde, e Rui Pires de Vila Verde, na ocasião em que foram

Operação corrente :

Imagem: Index das Gavetas Vol.1 - F.1

1
2
3
4
5
6
7
8
9
1
2

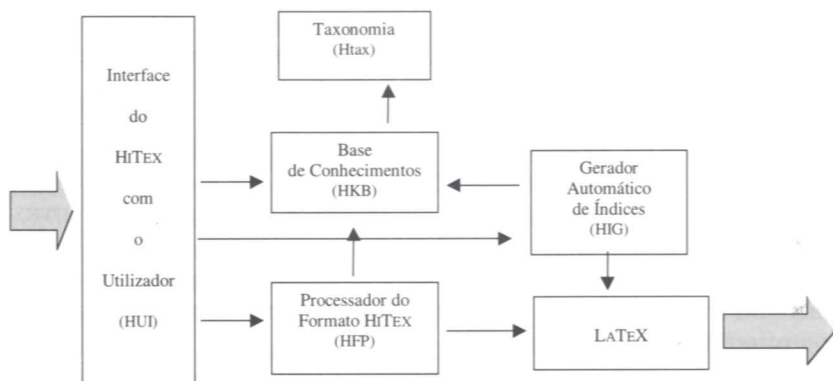
338

1
2
3
4
5
6
7
8
9
1
2

Extracção das Festas que os Cavalheiros de Braga fizeram aos Senhores Marquês de Aronches, Príncipe Senescal, Marquês de Vila Verde, e Rui Pires de Vila Verde, na ocasião em que foram

Figura 2 – Ecrã após ter recebido a imagem do índice das Gavetas do Cabido

Após a explanação da interface com o utilizador poder-se-á fazer a seguinte análise de *como* é o motor do sistema HITEX.



A figura representa o bloco estruturado do actual sistema HITEX. O bloco de Interface do HITEX com o Utilizador (HUI) é responsável pela gestão da interacção com o utilizador. Grande parte dessa interacção tem a ver com a criação e manutenção da Base de Conhecimentos (HKB) e correspondente taxonomia (HTAX) onde, como se viu, toda a informação é registada.

Os outros blocos estruturais do HITEX têm a ver com o tratamento de fontes (transcrição documental). HFP é o Processador do Formato HITEX atrás referido, produzindo código. Finalmente, o bloco Gerador Automático de Índices (HIG) pode ser activado para se gerarem automaticamente os índices toponímico, cronológico, etc. Mais uma vez, o texto dos índices é produzido em LATEX, e «colado» – se pretendido – ao texto já processado, para tudo ser impresso.

Em 1994 o projecto HITEX chegou ao fim da sua segunda fase de desenvolvimento após a construção de um protótipo escrito em SMALLTALK e C. Embora a interface com o utilizador esteja bastante mais aperfeiçoada do que a de um protótipo anterior em XMETOO, actualmente, com as novas tecnologias informáticas com grande incidência da interface gráfica, torna-se indispensável a readaptação do projecto, tanto no que diz respeito ao tratamento das imagens digitalizadas como no tamanho da base de conhecimentos.

Uma terceira fase do projecto está actualmente em desenvolvimento após o teste extensivo e avaliação pela equipa de historiadores do projecto.

Prevê-se a necessidade de aumentar a capacidade de dedução automática do sistema, assim como melhoramentos na produção de

índices textuais (a geração automática de linguagem natural requerá investimento adicional em linguística computacional).

Uma novidade desta integração consistirá na aplicação sistemática de um cálculo de implementação, desenvolvido independentemente pelo autor, com o objectivo de colmatar o desafio implementacional do HITEX.

Em 1995 desenvolveu-se o projecto HINET de divulgação do Arquivo Distrital de Braga na Internet, através de protocolo de colaboração com o Centro de Comunicações da Universidade do Minho e a PORTUGAL TELECOM, que visava a interligação do HITEX com a Internet.

Este projecto impulsionou o Arquivo Distrital de Braga a nível nacional e internacional, levando a um novo envolvimento no recente projecto GeIRA.

GeIRA: um sistema para divulgação cultural e científica

Duas grandes linhas orientadoras motivaram o nascimento do projecto GeIRA: a sociedade de informação, e a realidade social, económica e cultural da Euro-região Norte de Portugal – Galiza. Nesta perspectiva, o GeIRA é a ciência e tecnologia ao serviço da sociedade, afirmando-se como um sistema de difusão de informação apoiado na Internet – advento da sociedade de informação – e em outros meios de difusão de informação multimédia.

O projecto GeIRA arrancou em Janeiro de 1997 com financiamento atribuído pelo programa INTERREG II, o qual permitiu viabilizar, na zona de intervenção Portuguesa (Norte de Portugal), uma parte significativa da instalação de uma infra-estrutura distribuída de serviços, apoiada em equipas pluridisciplinares de investigadores de duas Universidades do Norte de Portugal: a Universidade do Minho (UM) e a Universidade de Trás-os-Montes e Alto Douro (UTAD).

O GeIRA é um projecto de desenvolvimento de serviços de informação multimédia e interactivos relativos ao património Português, sendo a área geográfica seleccionada para o seu arranque o Norte de Portugal. As principais vertentes de actuação do GeIRA são as seguintes:

- divulgar o potencial científico e tecnológico da região, em termos de pessoas, projectos de ensino e de I&D, e Instituições, e em especial ao nível das Instituições de Ensino Superior;

- utilizar as ciências – da informação e da comunicação – e as tecnologias – da multimédia e da interactividade – na valorização do património cultural e no estímulo à protecção e conservação do ambiente.

Os formatos de divulgação multimédia previstos no projecto baseiam-se essencialmente na utilização de servidores Web na Internet e na produção de CD-ROM's. Os servidores de informação na Internet são baseados em bases de dados integradas ou em cooperação, que suportarão formas de pesquisa múltipla por toda a informação existente no GeIRA, incluindo a pesquisa por texto livre, por intervalos de tempo, ou ainda por coordenadas geográficas (informação geo-referenciada).

No sentido de garantir adequados níveis de qualidade na execução do projecto, este foi estruturado em várias áreas de intervenção temática – de acordo com os objectivos e com as necessidades de funcionamento – nomeadamente Ciência e Tecnologia, Arqueologia, Arquivos, Bibliotecas, Museus e Natura.

GeIRA-Arquivos, HITEX e o futuro...

No âmbito do projecto GeIRA, o Arquivo Distrital de Braga assumiu a organização da equipa de trabalho da Área de Intervenção GeIRA-Arquivos na Universidade do Minho, através de protocolo com a Comissão Executiva do GeIRA.

Partindo das infra-estruturas, experiência e programas de pesquisa iniciados com o projecto HITEX e já utilizados no projecto HINET, o Arquivo Distrital de Braga propôs-se rentabilizar ainda mais o esforço dispendido, sobretudo, na caracterização e estruturação das bases de dados de informação arquivística, visando a concepção de normas nacionais de estruturação da informação via Internet e CD-ROM.

Um dos objectivos principais deste projecto que visa rentabilizar o estudo efectuado pela investigação anterior relacionada com o projecto HITEX é a concepção do sistema «HITEX++», uma nova versão para funcionamento em suporte www, como projecto de estágio de licenciatura.

Para além de:

- conceber um sistema de reconhecimento e cruzamento de informação retida nas diversas bases de dados já construídas

no ADB e/ou Arquivos Municipais, com os quais o ADB tem protocolo;

- conceber uma forma de enriquecer toda a informação disponível através da associação de fac-símiles da documentação microfilmada;
- elaborar um sistema de *compatibilização* da informação estruturada já existente com o acesso via Internet e/ou CD-ROM.

Após um primeiro ano de trabalho poder-se-á fazer um balanço positivo, tendo sido feita a reestruturação da informação arquivística existente sobre o Arquivo Distrital de Braga, tendo-se criado uma identidade própria www.adb.pt e novos serviços na Internet, com especial relevo na edição gráfica de livros digitais (o livro *Memórias Particulares* de José Inácio Peixoto editado em 1992 pelo ADB) e no tratamento de imagens em microfimes.

No que diz respeito a este último ponto – o tratamento de imagens em microfimes – a Área de Intervenção GeIRA-Arquivos adquiriu uma máquina digitalizadora de microfimes – ABR 3000D da BELL+HOWELL – com a qual prevê a digitalização integral dos microfimes do Registo Paroquial e de outros.

Estas imagens serão futuramente integradas no sistema «HITEX++» para pesquisa via Internet e CD-ROM.

BIBLIOGRAFIA

- AMADO, J. P.; CARDOSO, J. C.; NEVES, A. M.
Heródoto 1.0 Estação de Trabalho Informática em História e Arquivística. Lisboa: Fac. de Ciências Sociais e Humanas da U. N., 1990.
- ANDRADE, Pedro Luís
Implementação da Base de Conhecimento do HITEX em SMALLTALK. Braga: U. Minho, 1994. Relatório de Estágio, LESI.
- BARROCA, L.
Object-oriented Database Design in Archaeology. [Braga]: Dept. of Computer Science, University of Minho, 1991.
- GULAMHUSSEN, Asha
Desenvolvimento de uma base de dados relacional para o projecto HITEX. Braga: U. Minho, 1993. Relatório de Estágio, LMCC.
- MOREIRA, Clara Sofia
HITEX: A Integração e Teste do Sistema. Braga: U. Minho, 1994. Relatório de Estágio, LMCC.
- OLIVEIRA, J. N.; ARAÚJO, A. S.; SILVA, A. M.
«Historical Records Processing in the HITEX System». To appear in *Proc. of AHC'91 6th International Conference of the Association for History & Computing*. Odense, Denmark, Aug. 28-30, 1991.

OLIVEIRA, José Nuno

«HITEX: um sistema em desenvolvimento para historiadores e arquivistas», *Forum* (15/16), Braga, 1994, pp. 123-137.

Projecto HITEX. 4.º Relatório de progresso. Arquivo Distrital de Braga/Universidade do Minho, Agosto 1993.

«A Reification Calculus for Model-Oriented Software Specification», *Formal Aspects of Computing*, V. 2, (1-23) Springer-Verlag, 1990.

PEREIRA, António Manuel

HITEX: Um Sistema para Transcrição Documental em Larga Escala. Braga: U. Minho, 1991. Relatório de Estágio, LESI.

THALLER, M.

The Need for Standards: Data Modelling and Exchange in Modelling Historical Data: Towards a Standard for Encoding and Exchanging Machine Readable Texts, 1-18, Greenstein D.I. (Ed.), Max-Planck-Institut für Geschichte, St. Katharinen, 1991.