

Arquivos de vídeo e a Sociedade de Informação

Artur Pimenta Alves
Prof. Catedrático da FEUP
Director do INESC Porto
palves@inescporto.pt

1. Introdução
2. De Armazens a arquivos
3. Os metadados e o modelo de exploração
4. Actividade Relevante dos últimos 20 anos
5. VIDION
6. Lições
7. Arquivos de TV na Internet
8. Soluções

- Os programas de tv e vídeo são um património cultural importante mas, apesar disso, boa parte da produção actual não é preservada
- Segundo estimativa recente no mundo produzem-se 30 milhões de horas de programação nova de televisão em cada ano mas destas só uma pequeníssima parte é conservada (1).
- A redução do custo de armazenamento verificado nos últimos anos, após a digitalização, permite alterar esta situação.

- Apesar disso só na europa estima-se que existam já cerca de 100 milhões de horas de TV em arquivo.
- Este acervo contem informação cultural, histórica, sociológica, antropológica, educativa - para mencionar algumas áreas de interesse - muito relevante.
- Muito deste material está a ser preservado pelos seus próprios produtores ou, noutros casos, por entidades que o fazem com vista à sua exploração no mercado do audiovisual.

- Em Portugal assim sucede com as empresas de Televisão, mas muito particularmente com a RTP que possui um arquivo com mais de 300 mil horas de conteúdos, que tem vindo a preservar, tendo começado a digitalização dos conteúdos mais antigos.
- O arquivo da RTP foi transferido para o ANIM, mas os objectivos desta transferência não são ainda claros para mim, e poderão resumir-se apenas à partilha de espaços, mantendo-se a exploração a cuidado da RTP sem que haja o objectivo de facilitar o acesso para fins culturais ou de investigação, nomeadamente usando a Internet.

- Seria interessante adicionar todo este material à oferta de conteúdos acessível e pesquisável pelos estudiosos das várias disciplinas recorrendo à Internet. Seria uma grande acervo de conteúdos de banda larga em língua portuguesa.
- Em primeira avaliação, em termos mundiais, isso seria possível com uma verba pequena comparada com o custo anual das bibliotecas e das outras colecções actualmente preservadas e on-line (1, 2).
- Como atingir esse objectivo?

- Neste momento as barreiras que se opõem a que isto suceda não são técnicas embora ainda exista trabalho técnico e de investigação importante a fazer.
- Há que transformar os armazens de vídeo em arquivos catalogando-os por forma que seja possível a pesquisa de forma semelhante ao que sucede hoje com os documentos contendo texto ou imagens.
- Isso faz-se com metadados. Mas quais metadados? (3, 4, 5, 6)

- Uma primeira objecção ao progresso está na diversidade de objectivos que os metadados pretendem atingir e nos interesses das entidades detentoras dos arquivos que não são facilmente tornados compatíveis com este objectivo.
 - Por isso há várias normas e recomendações: DCMI, TVAnytime, MPEG-7, MPEG-21, SMPTE, EBU P/META,...
- os actuais operadores conservam os seus conteúdos para poderem proceder à sua reutilização na produção de programas.
 - Estudam questões como armazenamento de brutos, gestão de exhibições, gestão de direitos de propriedade, transacção de documentos com metadados e pouco se avançou na TV sobre IP ou na TV interactiva (MXF) que seriam as aplicações mais próximas da Internet.

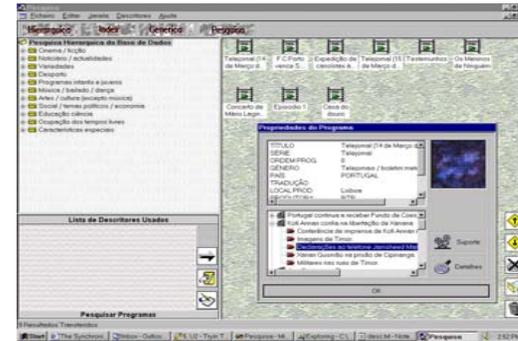
- Os actuais canais memória e os documentários históricos ilustram o que tem sido feito em termos de reutilização de conteúdos e para esse efeito é otimizada a catalogação.
- É claro que o modelo de negócio associado é completamente diferente daquele que poderá reger a sua disponibilização pela Internet para fins culturais ou de estudo.
- Mas catalogar vídeo é muito caro!
- A experiência do INESC Porto nesta área trouxe-me alguns ensinamentos e uma visão sobre estas questões.

- VIDION, com RTP, definição de metadados e ferramentas para catalogação (anotação) e para gestão de arquivos.
- ORBIT, com BBC, desenvolvendo uma solução para a produção em rede num estúdio digital suportando funções de captura, anotação, pesquisa, edição, armazenamento e mesmo “play-out”.
- Ferramentas de segmentação e anotação automática (geração de metadados)
- Sistemas de adaptação de conteúdos recorrendo a metadados (13 a 23)





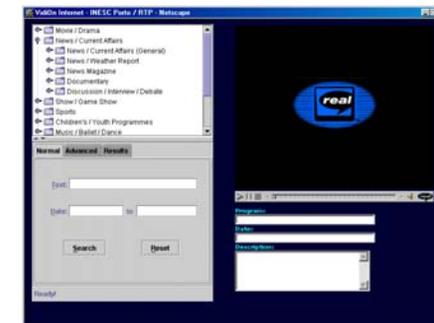
Anotação



Pesquisa



Edição



Pesquisa Internet

- Estes projectos permitiram-nos compreender a importância dos metadados, participar no desenvolvimento de algumas normas e compreender os diferentes âmbitos da sua aplicação.
- No entanto estes primeiros esforços não tiveram consequências. Não sucedeu apenas em Portugal - o problema está na confusão relativamente ao que se pretende fazer em termos de exploração do arquivo associado ao facto de a catalogação ser uma actividade muito exigente em recursos humanos.
 - Catalogação de notícias toma a um operador treinado 5h por cada hora

- Fizemos um sistema competitivo e com capacidade de permitir evolução e investigação em áreas essenciais. Ideias chave foram aproveitadas na infraestruturas desenvolvida com outros objectivos para a BBC.
 - Conceitos de metadata e de “browse quality”, bem como o uso de tecnologias IT genéricas e de baixo custo em broadcasting são hoje comuns, participamos na normalização de metadados com contribuições importantes nomeadamente no MPEG-7, 21 e no MXF.

- Mas

- Sistema inicial era rígido em termos de modelo de dados e por exigia customização pesada. Descobrimos soluções mas....
- Mais grave que tudo isso oferecíamos um conjunto de facilidades que o modelo de negócio da RTP não requeria naquela época pois a abertura do arquivo na Internet não fazia, ou faz ainda, parte da sua missão
- A RTP não se interessou por manter liderança na investigação no domínio e o projecto foi abandonado levando ao encerramento da empresa que tínhamos criado.

- Contudo, mais cedo ou mais tarde, teremos que ter acesso aos arquivos de TV e vídeo na Internet para fins de investigação ou de consulta.
 - [Doutoramento da Estrela Serrano lançou esse alerta nos media](#)
- Temos que encontrar soluções que defendam os interesses dos proprietários dos arquivos e resolver algumas questões legais associadas.
- Importa definir metadados e ferramentas de anotação e busca apropriados para que a experiencia não seja muito diferente dos motores de busca textual que actualmente utilizamos.

- Porque temos que adicionar metadados de descrição?
 - Existem metadados descritivos, administrativos e estruturais e estes tem que ser adequados à aplicação pretendida
 - O problema está em que uma máquina não pode procurar de forma eficaz uma cara de alguém mesmo que seja o tão conhecido BUSH? Ou cenas de por do sol? Ou cenas de fábricas de siderurgia? Ou cenas relativas a atentados? Ou cenas da vida doméstica? Ou cenas com cães ou pássaros? É mais fácil se esta informação for incluída nos metadados
 - Existe trabalho intenso nesta área no âmbito dos actuais motores de busca mas claramente que sem ferramentas de anotação que permitam a descrição não será útil podermos dispor de acesso a 100 milhões de horas de vídeo, existentes só na Europa!!!

- O que se faz é associar aos programas informação descrevendo-os detalhadamente de forma a que seja possível fazer a sua pesquisa usando os actuais motores de busca.
- Assim funcionava o VIDION, numa primeira fase apenas com anotação manual mas depois seriam desenvolvidas ferramentas automáticas para esse efeito.
 - Temos vindo a fazê-lo usando detecção de cortes de cena, detecção de movimento, detecção baseada na banda sonora, reconhecimento de fala (essencial suportar o Português que ainda não existe com qualidade), reconhecimento de orador,...

- Estado - Caso Francês
- Lei depósito Legal – toda produção original emitida em francês na rádio e TV é depositada e o INA arquiva. Há 7 ou 8 anos visitei-os e todo o material o material produzido era anotado manualmente recorrendo a anotadores (várias centenas) que preenchiam impressos em papel enquanto visionando “tapes” e anotando “time codes”. Tudo era posteriormente introduzido em bases de dados para consulta por estudantes e investigadores a partir de salas remotas (7).

- Estado - Caso Holandês
- Parlamento discute a possibilidade de aprovar lei que torne obrigatória a disponibilização aberta de todos os arquivos de televisão na Internet (8).

- Parcerias público-privado
- Dado que a digitalização e catalogação é um processo muito dispendioso têm sido realizados estudos com vista a descobrir que novos negócios são potenciados pelos arquivos na Net procurando detectar a possibilidade de fazer parcerias entre o estado e os privados interessados no processo, por forma a apoiar o Estado no financiamento da catalogação dos arquivos culturais e da televisão (Canadá) (9).

- Modelos abertos e colaborativos
- MIC (USA) e BBC são exemplos de programas que visam fornecer acesso livre a arquivos de vídeo para certos fins e estimulam a colaboração (up-load de vídeos e catalogação).
- O MIC definiu um modelo próprio de metadados e aceita conteúdos e conteúdos anotados em temas de acordo com regras que estimulam a colaboração e permitindo a consulta e a pesquisa (10).
- BBC (em colaboração com outros detentores de arquivos) anunciou algo semelhante para conteúdos seleccionados baseando-se nos direitos estabelecidos pela entidade designada “Creative Commons” (11, 12)

- Em Portugal é tempo de atacarmos o problema
- Temos que realizar estudos que demonstrem qual o caminho que devemos seguir.
- O INESC Porto está disponível para com parceiros que complementem as suas capacidades estudar soluções que se apliquem ao nosso país.
- Temos que estudar as experiências de países mais avançados de forma a propor formas de intervenção que nos permitam recuperar algum do atraso em que estamos.

- **Major Findings** - Educational access remains largely unattainable for a variety of reasons, including under-funding in public archives, lack of descriptive cataloguing and reference copies, copyright interests and very restrictive usage policies. Scholars best qualified to judge the long-term research value of television and video materials are generally not given ample opportunity to participate in decision making in public and corporate archives on what will be saved and made available.
- The network archives are focused on the daily production needs of broadcasters, constantly posing a danger that precious images so important to the collective memory of the American people will be lost, altered, or destroyed. Every group that has studied the selection of television for preservation has concluded that all news programs should be retained and preserved as aired.

- The **most devastating losses** have already occurred among news film and **videotape files of local television stations** across the United States. Even today local news tapes are rarely kept more than a week before they are recycled. About 20 states have no local television news collections in public archives, and very few libraries or archives take advantage of the right to make and retain off-air copies of daily newscasts.
- **The works of video artists and of independent video producers also face a precarious existence.**
- **Preservation** - Promotes the concept of a shared responsibility for the American television and video heritage, and calls for public and corporate archives to rationalize and coordinate their preservation programs to avoid unnecessary duplication and ensure that no significant portion of this heritage (held in collections throughout the nation) is endangered.

- **Access** - Encourages public and corporate archives to seek the advice and guidance of scholars and educators to establish appraisal standards and determine appropriate selection guidelines. Urges the identification of important television programs and coverage of events each year to encourage prompt availability in a public archives. Urges local television stations to work closely with advisory boards and local archives to halt further destruction of local news coverage. Recognizes the importance of video art and independent video production and calls for increased efforts to stimulate their collection.
- Urges the support of public policies that encourage the widespread dissemination of information through the Internet and other sources, and asks for a national union listing, a network of publicly shared databases, and a comprehensive catalog of American television programs by decade.
- Identifies steps to make it easier for scholars and educators to use television and video materials in their research, writing, and teaching, and calls for interested parties to intensify discussions

- **Funding** - Recommends the **establishment of an independent nonprofit organization in the private sector to raise funds for television and video preservation**, to recognize through an awards program individuals and organizations in this endeavor, and to keep television and video preservation at the forefront of the national archival agenda.
- Urges public archives to build a consensus around the principles of television and video preservation and make them understandable to funding organizations, which should then be more responsive to the needs of television and video archives.
- **Proposes discussions be held regarding possibility of two new avenues of funding: a dedicated sales tax and a share in future FCC auctions of broadcast spectrum.**
- Identifies the need for a documentary about the problems of television and video preservation aimed at general audiences and potential funders.
- Urges the inclusion of video art and independent video in all public awareness campaigns.
- **Developing an implementation plan is the next crucial step**

1. New Approaches to Television archiving, Jeff Ubois, First Monday, vol 10, nº 3, March 2005, http://firstmonday.org/issues/issue10_3/ubois/index.html
2. Television and video preservation 1997: A report on the current state of American television and video preservation: report of the Librarian of Congress, William T. Murphy, Library of Congress, ISBN 0-8444-0946-4
3. Digital Asset Management, David Austerberry, 2004, Focal Press, ISBN 0-240-51924-8
4. Descriptive Metadata for Television, Mike Cox, Ellen Mulder, Linda Tadic, 2006, Elsevier, ISBN 0-240-80730-8
5. Digital Video Archives: Managing Through Metadata, Howard Wactlar, Michael Christel, Carnegie Mellon University, <http://www.clic.org/pubs/reports/pub106/video.html>
6. Metadata Survey Report, Harry Salminen, CSC/Funet-TV, <http://tv.funet.fi/projekteja/tf-vvc/mdsurveyreport.jsp.en>
7. Institut National de l'Audiovisuel, <http://www.ina.fr/>

8. Dutch parliamentarians favour releasing public broadcast images into public domain, Joe Figueiredo, Digital Media News for Europe, 20/10/2004
9. Business Model Issues in the Development of Digital Cultural Content, Gerry Wall, First Monday, volume 8, nº 5, May 2003,
http://firstmonday.org/issues/issue8_5/wall/index.html
10. Moving Image Collection, <http://mic.imtc.gatech.edu/index.php>
11. bbc.co.uk Press Releases: BBC Creative Archive pioneers new approach to public access rights in digital age, 26th May, 2004
12. Creative commons, <http://creativecommons.org/>
13. Arquivos Digitais de Vídeo - Uma proposta de Arquitectura, Viana, Paula, Teixeira, Luís, Andrade, Teresa, Ribeiro, Hugo, Sequeira, João, Encontro Nacional de Bibliotecários, Arquivistas e Documentalistas, Aveiro, May 1998.
14. Vidion: An Online Digital Archive, Silva, Alexandre Ulisses, Viana, Paula, 6th DELOS Workshop, Tomar, Portugal, June 1998, pp. 67-70.
15. Viana, Paula, “Arquivos Digitais de Video: Uma proposta de Arquitectura”, 3as Jornadas RTP de Documentação e Arquivo, RTP, Lisboa, January 1999.

16. Richard Storey, Artur Alves, José Ruela, Luís Teixeira, Teresa Andrade, "The ATLANTIC news studio: reference model and field trial", Second European Conference on Multimedia Applications, Services and Techniques ECMAST'97, Milan, Italy, May 21-23, 1997, pp. 67-78.
17. Ferreira, Pedro, Cardoso, Pedro, Teixeira, Vítor Teixeira and Ruela, José, "An Open Architecture to Support Distributed Services in a Digital TV Studio", 107th AES Convention, New York, September 1999.
18. Cardoso, Pedro, ORBIT – Object Reconfigurable Broadcast using IT, DASE2001, NIST, Gaithersburg, MD, USA, June 19-20, 2001.
19. Santos, Ernesto, Veiga, Daniel, Middleware for the Broadcast Environment, DMMOS2001, Budapest, Hungary, June 18, 2001.
20. Teixeira, Vítor, Linking and transporting essence and Metadata over the production chain, 112th Audio Engineering Society Convention: Metadata Symposium, Munich, Germany, May 10-13, 2002.

21. Vitor Rodrigues, Mário Cordeiro, Paula Viana, José Ruela, "MXF Supporting the Integration of Media Applications and Broadcasting Products", Conftele'2003, 4th Conference on Telecommunications, June 18-20, 2003, Aveiro, Portugal.
22. M. T. Andrade, E. Santos, M. Tsakali and J. Livaditi, "Managing multimedia content and delivering services across multiple client platforms using XML", proceedings of the International Conference Mobile HCI 2002, Pisa, Italy (September 2002)
23. M.T. Andrade, P. F. Souto, P. M. Carvalho, L. Ciobanu, B. Feiten and I. Wolf, "Content Adaptation Decision to Enhance the Access to Networked Multimedia Content", proceedings of the WIAMIS 2006, Incheon, Korea, April 2006.