

## **A análise de programas televisivos com recurso a software.**

José Azevedo, Centro de Estudos das Tecnologias, Artes e Ciências da Comunicação da  
Universidade do Porto; Faculdade de Letras Universidade do Porto.

Cândido Pereira, Centro de Estudos das Tecnologias, Artes e Ciências da Comunicação  
da Universidade do Porto

A reconhecida importância da televisão na construção social da realidade das sociedades contemporâneas, a sua omnipresença na vida quotidiana dos indivíduos, e o crescimento global da produção audiovisual, tem levado a um aumento da investigação na área da comunicação das imagens em movimento.

Por outro lado, a convergência e aumento de capacidade do hardware permitiram que grandes quantidades de material audiovisual possam ser alvo de investigação com relativamente poucos recursos. O que ainda tem faltado é um conjunto de programas que permitam aos investigadores responder ao interesse e capacidade já instaladas.

Assim, nesta comunicação apresentamos uma revisão geral dos diferentes tipos de software que têm sido desenvolvidos para investigação do audiovisual, discutindo as suas potencialidades e limitações.

Gostaríamos de começar por referir que esta área está ainda em claro período de desenvolvimento, não sendo certo qual o tipo de software que se tornará dominante neste campo. Contudo, pode-se desde já dizer que o software desenvolvido se estrutura em dois grandes grupos. Um primeiro conjunto de programas procura essencialmente uma forma **automática** de organizar e pesquisar em grandes bases de dados. Um segundo grupo de programas, procura essencialmente possibilitar a codificação manual de material vídeo de forma a realizar estudos qualitativos mais aprofundados ou a simplesmente organizar conteúdos de uma forma mais personalizada.

Desta forma, optamos por estruturar a comunicação em torno destes dois grandes tipos de software.

### 1) Software para pesquisas automatizadas: da geração Google à geração Voogole

Hoje é possível utilizar sistemas altamente automatizados que simultaneamente com a gravação procedem a uma indexação de emissões de televisão ou outro material videográfico.

Esta indexação automática pode ser feita com base em três tipos de dados:

- 1º na informação contida no teletexto e que é difundida em sincronia com a emissão
- 2º em sistemas de reconhecimento do áudio que identificam silêncios, música e transcrevem o discurso
- 3º em sistemas de análise do conteúdo da imagem

### Indexação de vídeo baseada no teletexto

As emissões de televisão no sistema NTSF e PAL permitem a difusão de informação textual embebida no sinal vídeo. Esta possibilidade foi (e ainda é) explorada pelas estações de TV para auto-promoção e mais recentemente para traduzir a informação áudio fazendo-a assim chegar aos deficientes auditivos. Este “texto” (closed captioning) contém a informação dos vários elementos da pista de som: as vozes on e off e a identificação dos falantes, descrição do som ambiente e música. O formato VHS não permite reter esta informação que pode ser no entanto gravada no formato S-VHS ou mais eficazmente num formato digital. Estes dados constituem um manancial óptimo para a indexação do material vídeo, uma vez que podem ser tratados pelas mesmas ferramentas poderosíssimas que hoje possuímos para as bases de dados textuais.



Símbolo indicativo de que o programa é acompanhado por closed captioning

Infelizmente na televisão portuguesa esta informação ainda é rara e não se perspectiva uma aposta nesta área. Nos Estados Unidos a Telecommunications Act em

1996 estabeleceu um calendário para a progressiva inclusão desta informação textual (closed captioning) na programação de televisão. Este calendário determina que em Janeiro de 2006 100% do novo material vídeo distribuído tem de conter esta legendagem (informação disponível em [http://www.healthyhearing.com/library/news\\_content.asp?news\\_id=713](http://www.healthyhearing.com/library/news_content.asp?news_id=713)).

Deste modo os altos custos da produção das transcrições para texto da informação áudio são suportados pelos produtores e emissores dos programas possibilitando assim que os investigadores possam analisar grandes quantidades de material vídeo com recursos bastante modestos.

Greenberg & Lewenstein na comunicação que fizeram em Geneve 2001, dão conta da investigação que dirigiram servindo-se destes dados: Com apenas 3 desktop PCs equipados com sintonizador de TV e decodificador de teletexto monitorizaram as três grandes cadeias de TV americanas, ABC, NBC, CNN e ainda os late-night talk shows da CBS durante 6 meses, desde 1 de Agosto de 1999 até 31 de Janeiro de 2000. O objectivo era recolher referências ao bug do ano 2000 para posterior análise. O software usado gravava a emissão em blocos de 15 minutos. A informação de teletexto correspondente a cada bloco era pesquisada automaticamente no sentido de se detectar expressões chave relacionadas com o bug; se eram encontradas referências o bloco de vídeo era preservado se não era apagado. Ao fim dos seis meses os investigadores tinham reunido 590 horas de vídeo contendo 2363 referências independentes sobre o Y2K.

### **Sistemas de reconhecimento do áudio (audio mining)**

Esta tecnologia tem como objectivo a transcrição automática (por software informático) do discurso natural e descrição de outro tipo informação auditiva contida em documentos multimédia (emissões de rádio, televisão, comunicações telefónicas ou na Internet). Existem vários agentes interessados na utilização desta tecnologia:

- Os serviços de segurança de estado por motivos óbvios;
- Os produtores de vídeo porque poderiam assim acrescentar o closed captioning aos seus produtos poupando nos salários aos transcritores humanos;

– Os investigadores porque poderiam indexar assim muito do material videográfico não transcrito.

Considerada várias vezes “The next big thing” esta tecnologia ainda não atingiu, no entanto, a massa crítica que permita a banalização de aplicações como por exemplo acontece com o OCR (reconhecimento óptico de caracteres)<sup>1</sup>. Isto apesar de existirem já sistemas muito potentes e precisos.

Trazemos aqui como exemplo desta tecnologia o podscope ([www.podscope.com](http://www.podscope.com)).



Podscope is powered by [TVEyes](http://www.tveyes.com) - © 2005

#### Interface do podscope

O podscope trata-se de um motor de pesquisa a exemplo do google. A grande diferença está no facto do podscope trabalhar com publicações multimédia na net que são indexadas por um sistema de audio mining para língua inglesa. O sistema de indexação na base do podscope indexa podcasts e vlogs (vídeo blogs) que desde 2004 têm proliferado na net. Este motor é apenas um cartão de visita da TVEyes ([www.tveyes.com](http://www.tveyes.com)) que oferece (vende) indexação de emissões televisivas por este processo. O podscope pode ser o embrião da geração Voogole (Vê de vídeo); imagine-se as emissões de TV em servidores acessíveis na net e um motor de pesquisa com a

---

<sup>1</sup> Uma das limitações à expansão desta tecnologia é o facto de ser idioma-dependente. Existem vários sistemas para língua inglesa e curiosamente para língua árabe mas para o mercado americano, claro.

mesma eficácia que hoje tem o google a indexar e a devolver resultados numa base de imprensa escrita.

### **Sistemas de análise do conteúdo da imagem**

Os sistemas de que falamos até agora centram-se na informação áudio o que pode ser completamente suficiente para muitas situações em que a imagem está subordinada ao áudio, no entanto são ineficazes a indexar vídeos em que a palavra está subordinada pela imagem. Tal como o sistema de closed captioning leva a televisão aos deficientes auditivos podemos pensar num sistema de descrição das imagens capaz de aproximar os invisuais a este meio. Em termos computacionais a análise da imagem levanta obstáculos muito maiores que o áudio; como era de esperar é bem mais fácil ensinar os computadores a ouvir do que a ver. A investigação nesta área foi muito intensa na segunda metade dos anos noventa. Nessa altura abriram-se os caminhos que hoje ainda se estendem. A velocidade de progressão é condicionada essencialmente pela capacidade computacional mas também pressão (ou falta dela) dos utilizadores.

A análise do conteúdo das imagens vídeo implica a segmentação das peças (em unidades superiores à frame, claro) em elementos coerentes em termos visuais. Para isto os softwares usam parâmetros como a cor, textura, forma, posição e movimento. O resultado é uma segmentação em planos (shots). Estes constituem então as unidades de análise características destes sistemas. Os planos têm coerência sintáctica mas falta-lhe valor semântico importante para a classificação. Atendendo apenas às características das imagens é muitas vezes impossível agrupar os planos em cenas que já são unidades com semântica significativa. Assim muitos sistemas combinam as técnicas de análise de imagem com as outras tecnologias já aqui abordadas no sentido de conseguir uma mais poderosa indexação do material e consequentemente melhores resultados nos processos de pesquisa nas bases. Para a inquirição das bases produzidas por estas tecnologias não é suficiente o sistema da “palavra chave” daí que existam também métodos de pesquisa baseados na comparação de características primárias das imagens (por exemplo o histograma).

O exemplo que trazemos aqui é o sistema Marvel que foi desenvolvido pela IBM com a última edição em 2004. Este sistema é capaz de indexar enormes quantidades de vídeo

aplicando tecnologias de áudio mining, extracção de informação do teletexto e análise do conteúdo das imagens. São definidos segmentos correspondentes a planos e cada um é etiquetado com os resultados das técnicas mencionadas (este processo ocorre em tempo real). Ainda é possível acrescentar dados resultantes de observação humana. Simultaneamente com cada etiqueta é atribuído um valor numérico que corresponde ao grau de confiança desse “label”. Por exemplo, o software encontra um rosto num segmento e o algoritmo regista isso juntamente com 0,9 que significa o grau de confiança para a existência de um rosto nesse segmento é de 90%.

A inquirição da base pode ser feita por palavras-chave, por comparação com padrões escolhidos ou navegando pelas categorias semânticas definidas.

Podemos testar este sistema na Internet através do endereço <http://mp7.watson.ibm.com/marvel/>



Interface inicial do sistema de demonstração do MARVEL

## 2) Software para pesquisas não-automatizadas

Tal como referimos no ponto anterior a análise do conteúdo das imagens vídeo implica a segmentação das peças em elementos coerentes em termos visuais. No caso da maioria das investigações essas segmentações obedecem a critérios que são fundamentalmente

conceptuais e logo torna-se impossível atendendo apenas às características das imagens, agrupá-las de forma automática em categorias. Desta forma, o investigador deve proceder, tal como quando trabalha com texto, a uma análise de conteúdo manual, com os respectivos passos de codificação e segmentação de segmentos de imagem.

A codificação é essencialmente uma forma de indexação ou de identificação de categorias dos dados. Uma categoria é uma dimensão utilizada para classificar dados (neste caso, imagens). O objectivo da codificação é de facilitar a recolha de segmentos de dados (imagens) por categoria analítica. Utilizada desta forma a codificação simplifica e reduz os dados (Coffey & Atkinson, 1996; Miles & Huberman, 1994). Coffey and Atkinson (1996) mostram que “ A natureza dos dados qualitativos significa que dados relativos a um determinado tópico não estão reunidos exactamente no mesmo lugar de cada entrevista<sup>2</sup> (e as notas de campo tem uma organização ainda menos previsível). A capacidade para localizar extractos de dados que, pelo menos aparentemente, são “acerca” da mesma coisa é um aspecto valioso da gestão de dados” (p.35).

Organização dos dados, desenvolvimento de categorias, selecção de elementos específicos por categoria é um processo interpretativo que pode conduzir ao desenvolvimento de taxonomias ou outras formas de mais complexas de formulação de proposições que levem à compreensão mais profunda dos dados e à estruturação de quadros teóricos.

Para este tipo de necessidade de investigação vários programas têm sido desenvolvidos, exemplificamos nesta comunicação o 4vdo.

---

<sup>2</sup> A lógica do procedimento é idêntica no caso de uma entrevista ou no caso de um documentário

O 4vdo Annotator<sup>3</sup> é uma aplicação ainda em desenvolvimento por uma equipa de programadores do INESC Porto. Este software destina-se à criação e gestão de bases de dados vídeo. As anotações dos vídeos são feitas por mão humana. Está em desenvolvimento uma rotina que permite a segmentação automática das peças segundo os planos.

O programa permite:

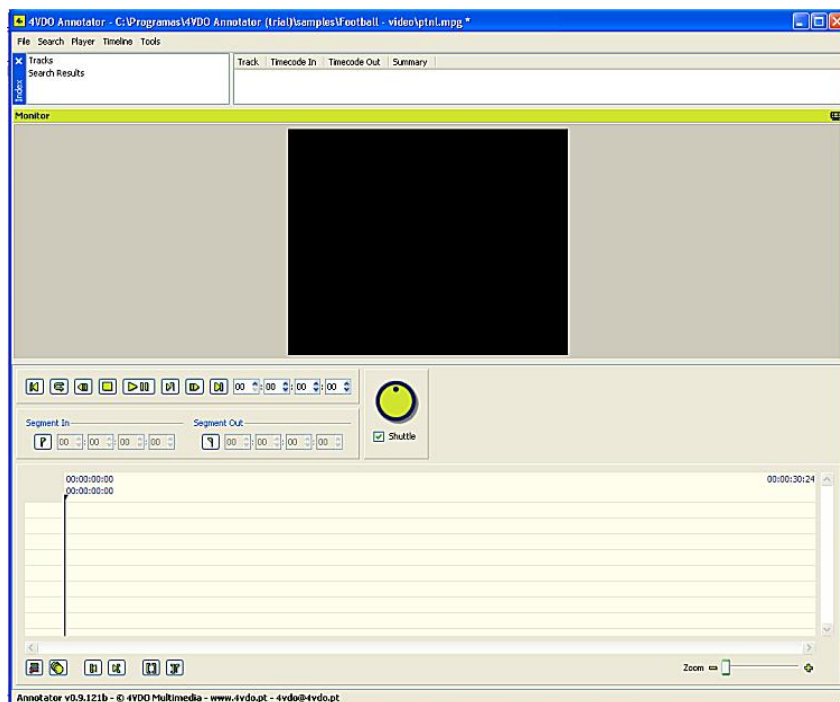
- 1 – Definir e estruturar os campos da base;
- 2 – Anotar as peças de vídeo na sua globalidade;
- 3 – Definir segmentos nas peças e anotá-los;
- 4 – Inquirir a base de vídeos mediante critérios definidos pelos utilizadores.

Os dados são guardados num formato XML seguindo a norma MPEG7. Este aspecto é de crucial importância porque vai permitir uma fácil integração das bases produzidas com outras e ainda a utilização de motores de pesquisa externos.

---

<sup>3</sup> Site: [www.4vdo.pt](http://www.4vdo.pt); Empresa: 4VDO - Sistemas e Serviços Multimédia, S.A.  
Propriedade e Copyrights: da empresa; Morada: Rua Dr. Roberto Frias, 378 4200 - 465 Porto , Portugal  
Contacto: [4vdo@4vdo.pt](mailto:4vdo@4vdo.pt)

## Exemplificação do trabalho com o 4vdo Annotator



O interface inicial do 4vdo annotator

### Iniciar um projecto

1º-Escolher o vídeo que deve estar na pasta onde o projecto será guardado.

2º-O título, autor e data são propriedades do projecto e poderão ser modificadas posteriormente.

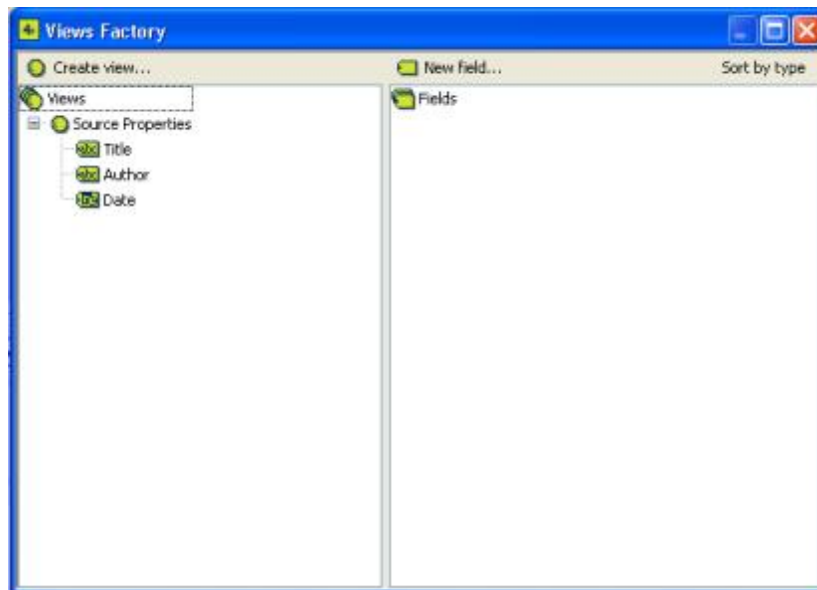
3º-Definir campos de identificação e caracterização geral da peça: Escolher “Tools → Views Factory...”



Para definir um campo começa-se por atribuir um nome e definir o tipo. No caso dos campos “Vocabulary Terms” define-se também os termos possíveis para o campo.

Estes campos devem ser colocados na “view” “Source Properties”. Para isso basta arrastar o campo para a sua posição nesta “view”. Aqui ainda se pode fazer subir ou descer a posição e definir o número de linhas que são apresentadas. Para aceder a estas possibilidades, faz-se botão direito do rato sobre o campo.

4º-Definir os campos e associá-los em “views”. Uma “view” deve conter um conjunto de campos que possam caracterizar um mesmo segmento do vídeo e que versem sobre o mesmo aspecto. Por exemplo pode definir-se uma “view” sobre aspectos relativos à imagem que podia conter um campo sobre o tipo de plano, outro sobre o ângulo, etc.



interface para criar a estrutura da base de dados

5º-Criar as “tracks”.



As “tracks” não são mais do que as “views” na “timeline”, isto é, a cada pista corresponde um conjunto de campos que tomarão valores nos segmentos que forem definidos nessa pista.

## **Anotar um vídeo**

1º-Preencher/ver as “source properties” escolhe-se “File → Properties”. Estas propriedades não estão dependentes dos segmentos que se venham a criar.

2º-Definir segmentos nas diferentes pistas. Começa-se por seleccionar a pista onde se quer o segmento, marca-se o ponto inicial , depois o final  e por fim define-se o segmento . Pode-se criar um segmento usando o botão  que fecha um segmento, define-o e marca o início do próximo.

Para definir o mesmo segmento noutra pista basta seleccionar o segmento, depois selecciona-se a pista e por fim usa-se o botão  para fazer o segmento nessa pista.

(O 4VDO terá uma possibilidade de criação de segmentos automaticamente por detecção de planos. Pode ajustar-se a “sensibilidade” à mudança que desencadeia o corte.)

3º-Introduzir os valores nos campos. A cada pista corresponde uma janela onde estão dispostos os campos. Aí pode-se proceder ao seu preenchimento e visualização. Estas janelas podem ser abertas e fechadas conforme as conveniências.

## **Inquirição da base de dados**

1º-O comando “Find” (Search → Find) permite encontrar os segmentos que contém a referência (conteúdo de um campo) pedida.

2º-O comando “Advanced Search” (Search → Advanced Search) permite fazer pesquisas mais criteriosas  
mas neste momento a implementação é pouco prática e pouco clara.

3º-“Search In Files” (Search → Search In Files) permite fazer o mesmo tipo de pesquisas mas agora pesquisa é feita em todos os projectos que se encontram numa pasta e respectivas sub-pastas.

## Conclusão

- Apesar de hoje ainda ser mais difícil analisar informação televisiva do que imprensa escrita, já existem ferramentas que tornam essa tarefa realizável, abrindo-se-se campos muito vastos para a investigação.
- Estão hoje definidas normas internacionais estáveis para a estruturação da informação multimédia, o MPEG7, pelo que a preferência dos investigadores deve ir para as ferramentas que usem este standard.
- Existem nichos nestas tecnologias que têm de ser desenvolvidos de acordo com especificidades linguísticas, logo também aqui existem oportunidades para os programadores e investigadores portugueses intervirem.

## Referências Bibliográficas

- Coffey, A. & Atkinson, P.** (1996). *Making sense of qualitative data*. Beverly Hills: Sage.
- Greenberg, P. & Lewenstein, B.** (2001) *Using Closed Captioning to Archive Science & Technology on TV: A case study*, PCST-6, Geneva, Switzerland
- Miles M. & Huberman, A.** (1994). *Qualitative data analysis: An expanded sourcebook*. Beverly Hills: Sage.
- Sebe, Nicu & Lew, Michael & Smeulders, A.**, (2003) Editorial, *Computer Vision and Image Understanding*, Volume 92, Issues 2-3, November-December, 141-146