



Proposta para Grupo de Trabalho

GTAaaS 2.0 – Acessibilidade como um Serviço com foco em pessoas com deficiência visual

Prof Dr Tiago Maritan Ugulino de Araújo
28 de Julho de 2015

1. Título

GTAaaS 2.0 – Acessibilidade como um Serviço com foco em Pessoas com Deficiência Visual

2. Coordenação Nome do coordenador: **Tiago Maritan Ugulino de Araújo**

Instituição: **LAVID/UFPB**

Lattes: <http://lattes.cnpq.br/6347743344931103>

Nome do coordenador adjunto: **Guido Lemos de Souza Filho**

Instituição: **LAVID/UFPB**

Lattes: <http://lattes.cnpq.br/6614550860293610>

3. Resumo

A audiodescrição (AD) é um importante recurso de acessibilidade para pessoas cegas. Ela consiste na narração de ações, de imagens, e de detalhes visuais, e é essencial para inclusão das pessoas cegas em experiências visuais como vídeos, aulas, televisão, cinema, etc. Neste projeto, estamos propondo a criação do GTAaaS 2.0 (Acessibilidade como um Serviço - versão 2), que tem como objetivo criar um serviço de geração

automática (ou semi-automática) de trilhas de AD em vídeos, tomando como base sua legenda e seu roteiro. Com isso, será possível tornar os serviços de vídeo da RNP também acessíveis para usuários e estudantes cegos. A proposta também envolverá o desenvolvimento de uma ferramenta de criação manual de roteiros de AD para vídeos, quando ele não puder ser gerado automaticamente.

4. Abstract

The audio description (AD) is an important accessibility feature for blind people. It consists in the description of actions, images, and visual details, and is essential for inclusion of blind people in visual experiences such as videos, classes, television, cinema, etc. In this project, we are proposing the creation of GTAaaS 2.0 (Accessibility as a Service - version 2), which aims to create an automatic generation service (or semi-automatic) of AD tracks in videos, from its subtitle and scripts. Thus, RNP's video services will be also accessible for blind users. The proposal will also involve the development of a tool for manual creation of AD scripts for videos, whether it can not be generated automatically.

5. Parcerias

ICPAC - Instituto dos Cegos da Paraíba Adalgisa Cunha

Contato: Valéria Cavalcanti Carvalho dos Santos

Descrição: O Instituto dos Cegos da Paraíba Adalgisa Cunha (ICPAC) é uma organização não governamental, de caráter filantrópico e assistencial, fundada pela Sra. Adalgisa Cunha, que teve como inspiração o Instituto Benjamin Constant (IBC) no Rio de Janeiro - RJ. O ICPAC é a primeira escola para pessoas com deficiência visual criada na Paraíba e atende crianças, jovens e adultos. A missão do ICPAC é a prestação de serviços voltados para a área da educação, assistência social, reabilitação visual e social, formação profissional e inclusão educacional, inserção no mercado de trabalho e plena inclusão social das pessoas com cegueira ou com baixa visão.

É importante ressaltar também que, caso o projeto seja aprovado, também pretendemos articular uma parceria com o **Instituto Benjamin Constant (IBC)**, um **Centro de Referência Nacional na Área de Deficiência Visual**, e instituição usuária da RNP. Por restrições de tempo, não foi possível articular essa parceria durante a elaboração deste projeto, mas é uma das metas durante a sua execução, a integração com esse centro, que é referência na área.

6. Duração do projeto

O projeto terá duração de 12 meses.

7. Sumário executivo

7.1. Justificativa

Uma grande carga de informação é veiculada através de meios audiovisuais. Em uma sociedade que apresenta um aumento no uso desses conteúdos como fonte de informação e entretenimento, as pessoas cegas ou com baixa visão podem ser excluídas por enfrentarem barreiras de acesso a esses conteúdos.

As pessoas com alguma deficiência visual possuem dificuldades para captação e compreensão de informações essenciais em experiências audiovisuais como aulas, vídeos, televisão, cinema, teatro, óperas, dentre outros. Em consequência disso, em muitas situações, elas acabam ficando excluídas do contexto sociocultural e intelectual da sociedade, uma vez que não são disponibilizados recursos de acessibilidade que lhes garantam a participação e o consumo desses conteúdos. Para garantir a acessibilidade a conteúdos, a preocupação não deve estar apenas na criação da informação. É preciso trabalhar a forma de apresentação, visualização e interação com o usuário. As dificuldades e limitações do usuário devem ser estudadas, a fim de propiciar meios de acesso à informação.

Segundo o censo demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2010, existem cerca de 35,7 milhões de pessoas com algum nível de deficiência visual no Brasil, o que representa cerca de 18,8% da população brasileira [5]. Dentre as deficiências investigadas, a deficiência visual apresenta a maior incidência na população, ocasionando, dessa forma, limitações no acesso às informações visuais, como televisão, cinema, vídeos, dentre outros. Além do aspecto social e cultural, a indústria audiovisual tem se tornado uma das mais importantes do mundo, pois, devido a convergência tecnológica, modernas infraestruturas de comunicação estão sendo projetadas em escala mundial.

Para reduzir essas barreiras de acesso, um recurso de acessibilidade bastante útil para as pessoas cegas é a audiodescrição (AD), que se apresenta como um meio de fornecer informações adicionais ao público. A audiodescrição é um recurso de acessibilidade desenvolvido para melhorar o acesso de pessoas, cegas ou não, a

essas informações visuais, permitindo que elas participem de experiências intrinsecamente visuais, como vídeos, através de descrições de imagens, narração de ações e detalhes visuais como figurinos, indicação de tempo e espaço [8]. Mais especificamente em vídeos, o objetivo da audiodescrição é a tradução dos elementos visuais que é inserida entre os diálogos, através de uma narração, de forma que não haja interferência nos efeitos sonoros do áudio original [1]. Essa tradução transforma toda a informação visual em palavras, ou seja, audiodescreve tudo o que é visto.

Para audiodescrever os vídeos, os roteiros de audiodescrição são geralmente produzidos de forma manual, o que demanda muito tempo. Em média, são necessárias cerca de 60 horas para um profissional descrever um vídeo com 2 horas de duração [7], por exemplo. Ademais, segundo pesquisas com empresas que prestam o serviço de audiodescrição, o custo para a criação de um roteiro de AD é de R\$ 40,00 por minuto de vídeo. No entanto, diversos serviços e plataformas de acesso a vídeos ainda não contam com este recurso de acessibilidade, e, em consequência disso, os deficientes visuais são exonerados de ter acesso à informação.

Na literatura científica, alguns trabalhos são direcionados para a aplicação da audiodescrição. No caso da audiodescrição específica para vídeos, alguns trabalhos abordam o uso de sintetizadores de voz na narração [2,6,9], a apresentação de

diferentes níveis de detalhamento de audiodescrição com versões resumidas e estendidas [2,3] e o uso de players próprios para facilitar a adição e consumo da audiodescrição [2,6]. Quanto à geração automática das descrições, pesquisas, em estágios iniciais, estão sendo feitas com o uso de técnicas de reconhecimento de elementos do vídeo [4] e de análise de roteiro cinematográfico sem vínculos com o tempo do vídeo [7]. Contudo, essas soluções apresentam limitações que ainda não atingem as necessidades de automatização da AD de forma satisfatória. Diante disso, a geração automática ou semi-automática da audiodescrição de vídeos, apresenta-se como uma área pouco explorada.

Para reduzir esses problemas, uma solução automática ou semi-automática para gerar trilhas de AD pode auxiliar nesta tarefa, tornando o processo de descrição mais eficiente, diminuindo o tempo investido e o custo associado. Dentro deste contexto, neste projeto, estamos propondo a criação do GTAaaS 2.0 (Acessibilidade como um Serviço - versão 2) que tem como objetivo desenvolver um serviço para geração automática de conteúdos acessíveis para pessoas cegas. A ideia é apresentar uma solução para a geração automática de AD de vídeos, na língua portuguesa, tomando como base o roteiro, legendas e/ou outros metadados relacionados ao vídeo. Quando o roteiro ou legenda não estiver disponível, uma ferramenta de auxílio para a criação semi-automática da audiodescrição deve ser utilizada. Dessa forma, é possível ter uma redução significativa dos custos de geração e produção de trilhas de AD, viabilizando o acesso a esses conteúdos para usuários cegos.

Uma vez que o projeto GT-AaaS aprovado pelo programa de GTs da RNP em 2011, por questões de tempo e recursos, endereçou apenas a acessibilidade para pessoas surdas, a ideia deste projeto é continuar explorando esse conceito, incorporando nele também a acessibilidade para pessoas cegas ou com baixa visão. Dessa forma, seria possível também incluir suporte a audiodescrição nos serviços de vídeo da RNP, como por exemplo o Video@RNP, Videoaula@RNP, Rede de Cinemas, ICD, RITU, dentre outros, tornando-os também acessíveis para pessoas com deficiência visual. Maiores detalhes sobre a proposta são apresentados na Seção 7.3.

É importante ressaltar que, de acordo com o Artigo 47 do Decreto 5.296 de 2 de Dezembro de 2004, é "*obrigatória a acessibilidade nos portais e sítios eletrônicos da administração pública na rede mundial de computadores (internet), para o uso das pessoas portadoras de deficiência visual*". Além disso, o Modelo de Acessibilidade em Governo Eletrônico (eMag), versão 3.1., reforça que "*vídeos que transmitem conteúdo visual que não está disponível na faixa de áudio devem possuir uma audiodescrição*" (Recomendação 5.3).

Para ajudar a compreender o potencial da solução proposta, um vídeo de demonstração foi elaborado usando um primeiro protótipo dessa solução desenvolvido pelo LAVID/UFPB: <http://bit.ly/1Ouh6xj>. O vídeo, separado em 3 partes, tenta mostrar, em sua primeira parte, a dificuldade das pessoas cegas ao assistir vídeos sem audiodescrição. Nas duas partes seguintes, é ilustrado como a audiodescrição pode enriquecer a experiência de uso dessas pessoas para assistir vídeos, auxiliar no processo de inclusão social e cultural e no acesso à informação.

7.2. Objetivos do Projeto

O objetivo geral deste projeto é o desenvolvimento de um serviço para geração automática trilhas de audiodescrição, reduzindo as barreiras de acesso a informação das pessoas cegas ou com baixa visão. O serviço proposto receberá vídeos legendados com roteiros (e possivelmente outros metadados do vídeo), gerando automaticamente

uma trilha de audiodescrição nesses vídeos e tornando-os, portanto, acessíveis. Quando a legenda e o roteiro do filme não estiver disponível, o serviço fornece uma ferramenta de suporte à criação manual da audiodescrição.

Para alcançar o objetivo geral foram traçados os seguintes objetivos específicos:

1. Com base em requisitos coletados com pessoas cegas, especificar e desenvolver um serviço para geração automática de roteiros de audiodescrição para vídeos, através da análise do roteiro do vídeo, da legenda e, possivelmente, de outros metadados relacionados ao vídeo;
2. Especificar e desenvolver uma ferramenta para apoio a criação manual da audiodescrição, utilizando como base requisitos coletados com profissionais na área de Audiodescrição;
3. Com base no roteiro de audiodescrição gerado de forma automática ou semiautomática, produzir uma trilha de Libras sincronizada com o vídeo principal, contando com o auxílio de ferramentas de síntese de voz;
4. Validar o protótipo desenvolvido com usuários surdos e profissionais audiodescritores.

7.3. Detalhamento do Projeto

Conforme mencionado na Seção 7.1, a proposta do projeto é continuar explorando o conceito de Acessibilidade como um Serviço, incorporando nele também a acessibilidade para pessoas cegas através da geração de trilhas de audiodescrição nos serviços de vídeo da RNP. Uma visão geral da solução, considerando a sua integração com os serviços de vídeo da RNP, é apresentada na Figura 1.

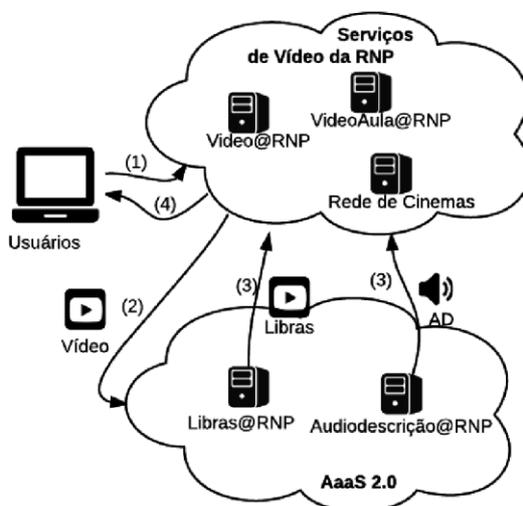


Figura 1. Visão esquemática do AaaS 2.0

De acordo com a Figura 1, o usuário ao submeter (ou acessar) um vídeo nos serviços de vídeo da RNP (Figura 1 (1)) pode optar por gerar também trilhas acessíveis usando o AaaS 2.0 (Figura 1 (2)). Caso o usuário seja surdo, o serviço Libras@RNP é invocado e gera automaticamente uma trilha de Libras para aquele conteúdo (Figura 1 (3)). Por outro lado, se o usuário for cego ou possuir algum nível de deficiência visual, o serviço proposto neste trabalho, denominado Audiodescrição@RNP, seria invocado (Figura 1 (3)) e geraria, de forma automática ou semi-automática, uma trilha de AD para aquele conteúdo (Figura 1 (4)), e tornando-o, portanto, também acessível para pessoas cegas. Uma visão mais detalhada do serviço Audiodescrição@RNP, proposto neste projeto, é apresentado na Figura 2.

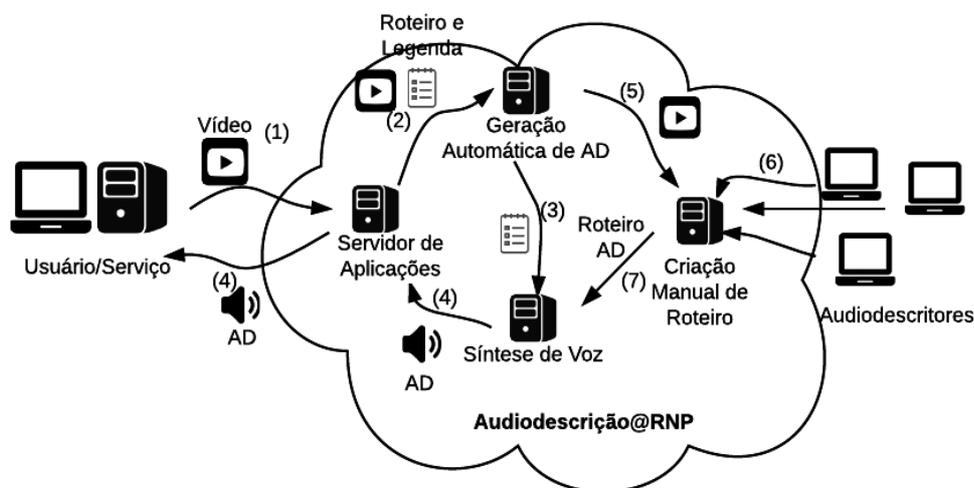


Figura 2. Visão esquemática do serviço proposto

De acordo com a Figura 2, o serviço proposto funcionará da seguinte forma. Inicialmente, um usuário ou um serviço submete um vídeo para o Audiodescrição@RNP (Figura 2 (1)). Esse vídeo é recebido por um servidor de aplicações que o encaminha para um (sub) serviço de Geração Automática de Audiodescrição (Figura 2 (2)). Caso o roteiro e a legenda desse vídeo estejam disponíveis, o serviço de Geração Automática de Audiodescrição, como o próprio nome indica, gera automaticamente um roteiro de AD contendo as descrições e as marcações de tempo e os encaminha para o (sub) serviço de Síntese de Voz (Figura 2 (3)). Este serviço recebe um roteiro de AD e gera uma trilha de AD sincronizada com o vídeo original. A trilha de AD é então encaminhada para o usuário/serviço cliente (Figura 2 (4)).

No entanto, caso o roteiro e a legenda do vídeo não estejam disponíveis, o vídeo é encaminhado para uma Ferramenta de Criação Manual de Roteiros de Audiodescrição (Figura 2 (5)). Nessa ferramenta, inspirada nos conceitos de computação por humanos e adaptada para geração de roteiros de AD, audiodescritores humanos podem, de forma colaborativa, criar um roteiro de AD para o vídeo (Figura 2 (6)). Esse roteiro é então enviado para o serviço de Síntese de Voz que repete o processo, isto é, converte o roteiro para uma trilha de AD sincronizada com o vídeo e o devolve esta trilha ao usuário.

Nas Subseções 7.3.1 e 7.3.2, serão apresentados alguns detalhes sobre o (sub) serviço de Geração Automática de Audiodescrição e a Ferramenta de Geração Manual de AD.

7.3.1. Serviço de Geração Automática de Audiodescrição

Conforme mencionado anteriormente, o serviço de geração automática de audiodescrição é responsável por gerar um roteiro de audiodescrição contendo os intervalos de tempo e descrições. A geração automática do roteiro de AD é feita a partir da análise do roteiro e da legenda do vídeo. A visão esquemática do serviço de geração automática do roteiro de AD é apresentada na Figura 3.

De acordo com a Figura 3, esse serviço funciona da seguinte forma. Inicialmente, o roteiro original do vídeo é analisado e seus principais elementos são extraídos, como títulos de cena, ações, personagens e outros. Em seguida, através da análise do vídeo e da legenda, o componente de Identificação de Gaps realiza a detecção dos intervalos

de tempo entre os diálogos, o que os caracterizam como possíveis intervalos sem falas, candidatos para futuras inserções de audiodescrição. Em seguida, o componente de Sumarização faz a extração das sentenças mais importantes do roteiro e, desta forma, resume o roteiro original, descartando as informações secundárias, consideradas menos importantes para a audiodescrição. Por fim, o componente de Geração de Roteiro de AD gera o roteiro de audiodescrição alocando as sentenças que foram extraídas na etapa de sumarização dentro dos *gaps* detectados na etapa de Identificação de *Gaps*. O resultado é o roteiro contendo as descrições das ações do vídeo qual, por sua vez, pode ser utilizado por narrador humano ou software de sintetização de voz na criação da audiodescrição.

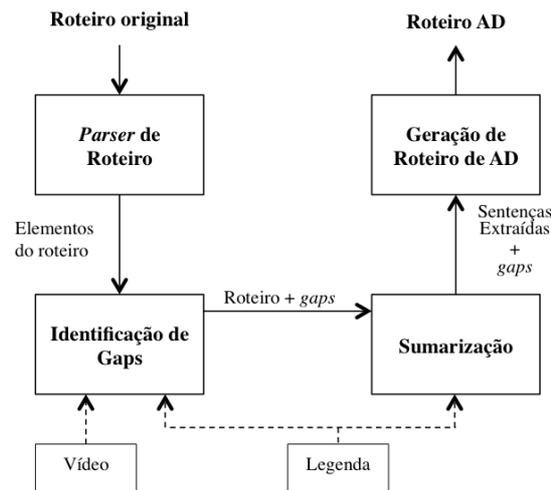


Figura 3. Visão Esquemática do Serviço de Geração Automática do Roteiro de AD

No (sub) serviço de Síntese de Voz, o roteiro de AD passa por um componente de síntese de voz que converte as sentenças do roteiro para voz (trilha de AD) e a sincroniza com o vídeo original.

7.3.3. Ferramenta de Criação Manual da audiodescrição

A Ferramenta de Apoio à criação manual da audiodescrição é uma ferramenta web colaborativa, inspirada nos conceitos de computação por humanos ou *crowdsourcing*, que é responsável por criar um ambiente e fornecer funcionalidades que auxiliem o profissional audiodescritor na criação do roteiro de AD. Este serviço poderá ser utilizado quando os elementos envolvidos na geração automática do roteiro de AD não estiverem disponíveis, como, por exemplo, o roteiro e a legenda. Alguns dos requisitos do serviço são os seguintes: (1) Criação de um ambiente de visualização e manipulação do vídeo, roteiro cinematográfico e legenda; (2) Ferramentas de identificação e marcação dos intervalos sem falas; (3) Criação e edição do roteiro de AD com opções de ajuste do texto das descrições, tempo e velocidade; (4) Gravação do áudio da narração durante a edição do roteiro ou uso de um sintetizador de voz para produzir a trilha de audiodescrição.

8. Recursos financeiros

8.1. Equipamentos e softwares

| Descrição | Quantidade |
|------------------------------------------------|------------|
| Desktop s/ monitor – Anexo II do Edital | 4 |
| Monitor LED 23.8" – Anexo II do Edital | 4 |
| Servidor – Anexo II do Edital | 1 |
| Licença de Software para Text-to- Speech Ivona | 1 |

9. Ambiente para testes do protótipo

Para validar o serviço apresentado nesta proposta, dois cenários de testes são sugeridos. O primeiro tem como objetivo utilizar uma parte dos recursos da infraestrutura do LAVID/UFPB. Neste cenário, o serviço de audiodescrição deve ser configurado e hospedado em um servidor dedicado nas instalações do LAVID para atender às requisições. O segundo cenário teste é direcionado para os usuários cegos da Fundação Nacional de Apoio ao Deficiente (FUNAD) e do Instituto dos Cegos da Paraíba Adalgisa Cunha (ICPAC). Para isso, sugerimos a elaboração de dois protótipos. A seguir, apresentamos a proposta de infra-estrutura para cada um desses protótipos.

Piloto 1: Geração de conteúdos acessíveis na infraestrutura do LAVID

Cenário: Utilizar um servidor de alto desempenho para executar o serviço usando a infraestrutura do LAVID.

Detalhamento: Desenvolver um protótipo do serviço e montar uma infraestrutura para executá-lo em um servidor disponível nas instalações do LAVID/UFPB. A infraestrutura proposta prevê o uso da Rede Ipê da RNP para o tráfego (*download* e *upload*) dos conteúdos multimídia. A **Figura 4** ilustra o cenário para implantação do protótipo.

Piloto 2: Geração de conteúdos acessíveis para usuários da FUNAD, ICPAC e IBC

Cenário: Utilizar um servidor de alta capacidade na infraestrutura do LAVID para executar o serviço e atender às requisições dos usuários cegos da FUNAD e do ICPAC.

Detalhamento: O objetivo do Piloto 2 consiste em realizar uma avaliação do protótipo desenvolvido com um grupo de usuários com deficiência visual do **Instituto dos Cegos da Paraíba Adalgisa Cunha (ICPAC)** e da Fundação Centro Integrado de Apoio a Deficiência (FUNAD) da Paraíba. O objetivo é avaliar o nível de compreensão e de satisfação dos usuários com os protótipos desenvolvidos. A infraestrutura proposta prevê o uso da Rede Ipê da RNP para o tráfego (*download* e *upload*) dos conteúdos multimídia.

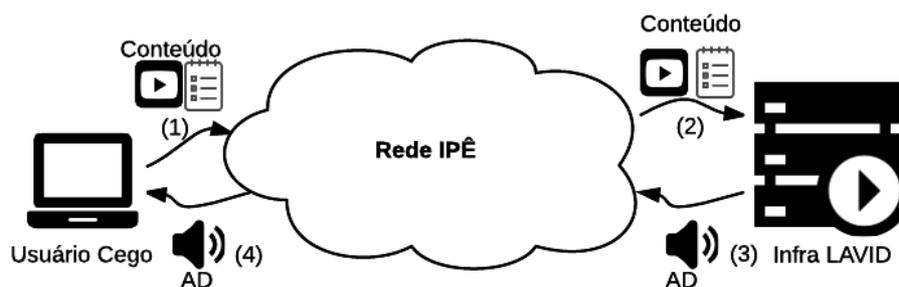


Figura 4. Visão esquemática da infra-estrutura do protótipo

Conforme mencionado nas Seções 5, caso ao longo do projeto, pretende-se também articular uma parceria com o **Instituto Benjamin Constant (IBC)**. Caso tenhamos êxito nessa tarefa, também pretende-se realizar uma avaliação com usuários dessa instituição.

10. Referências

- [1] BENECKE, B. **Audio-description**. In: GAMBIER, Y. (ed.), 2004. Meta. Volume 49, nº.1, p. 78-80.
- [2] CHAPDELAINE, C.; GAGNON, L. **Accessible Videodescription On-Demand**. ASSETS '09, Pennsylvania, p 221-222, oct. 2009.
- [3] ENCELLE, B.; BELDAME, M. O.; PRIÉ, Y. **Towards the usage of pauses in audio described videos**. 22nd International World Wide Web Conference, Rio de Janeiro, may 2013.
- [4] GAGNON, Langis et al. **Towards computer-vision software tools to increase production and accessibility of video description for people with vision loss**. Springer-Verlag, p 199-218, feb. 2009.
- [5] IBGE. Censo Demográfico 2010: Características gerais da população, religião e pessoas com deficiência. **IBGE- Instituto Brasileiro de Geografia e Estatística**, 2010. Disponível em: <<http://bit.ly/1S46pl4>>. Acesso em: 28/01/2014.
- [6] KOBAYASHI, M.; NAGANO, T.; FUKUDA, K.; TAKAGI, H. **Describing Online Videos with Text-to-Speech Narration**. W4A2010, Raleigh, apr. 2010
- [7] LAKRITZ, J.; SALWAY, A. **"The Semi-Automatic Generation of Audio Description from Screenplays"**, Dept. Of Computing Technical Report CS-06-05, 2002, University of Surrey.
- [8] NUNES, E. V.; MACHADO, F. O. e VANZIN, T. (2011). **Audiodescrição como Tecnologia Assistiva para o Acesso ao Conhecimento por Pessoas Cegas**. Ambiente virtual de aprendizagem inclusivo / organizadores: Vania Ribas Ulbricht, Tarcísio Vanzin e Vilma Villarouco. – 352 p. Florianópolis : Pandion, 2011.
- [9] O'CONNELL, T.; KOBAYASHI, M.; GOULD, B.; TAKAGI, H.; ASAKAWA, C. **Are Synthesized Video Descriptions Acceptable?** ASSETS '10, Florida, p. 163-170, oct. 2010.