



Proposta para Grupo de Trabalho

**GTAAA – Acessibilidade como um Serviço
(Accessibility as a Service)**

Prof Dr Guido Lemos de Souza Filho
18 de Agosto de 2011

1. Título

GTAAaS – Acessibilidade como um Serviço

2. Coordenação

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

Instituição: **LAVID/UFPB**

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

Nome do coordenador adjunto: **Tiago Maritan Ugulino de Araújo**

Instituição: **LAVID/UFPB**

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

3. Resumo

Pessoas com deficiência enfrentam sérias dificuldades para acessar informações. As tecnologias de informação e comunicação (TICs) quando são desenvolvidas dificilmente levam em conta os requisitos específicos destes usuários especiais. Neste projeto exploraremos o conceito de Acessibilidade como um Serviço, implementando e testando um serviço que torne possível o acesso a conteúdo digital para usuários especiais. A proposta é adicionar uma camada entre os sistemas legados e os usuários especiais, adaptando a apresentação dos conteúdos digitais para as necessidades destes usuários. Considerando as restrições de recurso e tempo do projeto, trataremos inicialmente de acessibilidade para usuários da RNP com deficiência auditiva, um grupo de usuários que apresenta maior dificuldade de comunicação devido à escassez de soluções técnicas adequadas para a Língua Brasileira de Sinais (LIBRAS).

4. Abstract

People with disabilities have serious difficulties to access information. The information and communication technologies when developed rarely take into account the specific requirements of these people. In this project we explore the concept of accessibility as a service by providing, implementing and testing a service that makes it possible to access digital content for disability users. The proposal is to add a layer between the legacy systems and disability users, addressing the presentation of digital content to their needs. Considering the resource and time constraints of the project, we will focus, initially, on RNP deaf users, a group of users that deals with difficulties of communication due to the lack of proper technical solutions for the Brazilian Sign Language.

5. Parcerias

CPqD – Centro de Pesquisa e Desenvolvimento em Telecomunicações

Contato: Lara Schibelsky Godoy Piccolo

CV Lattes: <http://lattes.cnpq.br/5053149056051004>

Descrição: O CPqD é uma instituição focada na inovação e tecnologias da informação e comunicação (TICs), e que tem como missão contribuir para a inclusão digital da sociedade brasileira. Ele ocupa uma posição de vanguarda tecnológica e mantém um dos maiores programas de pesquisa e desenvolvimento da América Latina. O portfólio do CPqD está alinhado com os temas de maior repercussão da atualidade: banda larga, nas telecomunicações; smart grid, no setor elétrico; banco do futuro, no setor financeiro; cidades digitais, na gestão pública.

6. Duração do projeto

O projeto terá duração de 12 meses.

7. Sumário executivo

7.1. Justificativa

Os surdos se comunicam naturalmente através de línguas gestuais denominadas línguas de sinais. Diferentemente das línguas orais que utilizam o som na comunicação, as línguas de sinais utilizam um canal visual, isto é, um conjunto de elementos lingüísticos manuais, corporais e faciais para articular os sinais [1]. O emissor constrói uma sentença a partir desses elementos lingüísticos e o receptor utiliza o sistema visual para compreender o que está sendo comunicado. Isso implica que para os surdos, as línguas orais, utilizadas cotidianamente pela maioria das pessoas e em praticamente todos os meios de comunicação, quando conhecidas, representam apenas “uma segunda língua” [2].

Em consequência disso, a maioria dos surdos têm muita dificuldade de compreender e se comunicar através de textos em línguas orais. Uma vez que essas línguas possuem uma grafia baseada em sons, muitos deles passam vários anos na escola e não conseguem aprender a ler e escrever na língua oral de seu país [3]. Essa dificuldade resulta em uma grande barreira para a comunicação com outras pessoas, o acesso a informações, a aquisição de conhecimentos, dentre outros.

Na literatura científica, alguns trabalhos estão direcionados para as necessidades comunicativas dos surdos [4] [5] [6] [7]. Esses trabalhos oferecem soluções tecnológicas para atividades cotidianas que permitem que pessoas com necessidades especiais assistam e compreendam televisão, interajam com outras pessoas ou escrevam uma carta. Por exemplo, o uso de legendas emotivas em filmes e programas de televisão [4] e o desenvolvimento de jogos para crianças surdas [5].

Contudo, a utilização de línguas de sinais em sistemas computacionais ainda é bastante limitada. A sua utilização em programas de TV, por exemplo, é ainda limitada aos dispositivos manuais, onde uma janela com um intérprete de língua de sinais é apresentada junto com o vídeo original do programa (*wipe*). Essa solução possui alguns problemas: (1) possui altos custos operacionais de geração e produção (câmeras, estúdio, equipe, etc.), (2) necessita de um intérprete em tempo integral e (3) pode ser incômodo para os telespectadores ouvintes, uma vez que essa janela ocupa boa parte do vídeo do programa.

Uma solução alternativa para esse problema seria gerar automaticamente a janela em língua de sinais, a partir da tradução automática e em tempo real do áudio ou da legenda do vídeo. Dentro desse contexto, estamos propondo a criação do GTAaS (Acessibilidade como um Serviço) que tem como objetivo desenvolver um serviço para geração automática de conteúdos acessíveis. A idéia é que usuários submetam vídeos ao serviço e ele gere automaticamente uma janela em línguas de sinais, mais especificamente, em Língua Brasileira de Sinais (LIBRAS), para esses vídeos, a partir do áudio ou da legenda do vídeo, tornando-os acessíveis para os surdos brasileiros. Dessa forma, tem-se uma redução significativa dos custos de geração e produção e uma redução da necessidade do intérprete de LIBRAS viabilizando o acesso a estes conteúdos para os usuários surdos.

Cabe ressaltar que a gramática de LIBRAS difere da Língua Portuguesa e de outras línguas de sinais. Por esse motivo é importante envolver o usuário surdo tanto no levantamento de requisitos quanto na validação da solução, de forma a garantir a melhor forma de apresentação e a inteligibilidade da informação gerada. A proposta é que o levantamento de requisito e os testes sejam realizados utilizando a infraestrutura do laboratório de usabilidade do CPqD.

7.2. Objetivos do Projeto

O objetivo geral desse projeto é o desenvolvimento de um serviço para geração automática de janelas em língua brasileira de sinais (LIBRAS). O serviço proposto receberá vídeos legendados e será capaz de gerar automaticamente janelas de LIBRAS para esses vídeos, tornando-os acessíveis. Os vídeos dos sinais de LIBRAS serão representados através de um agente animado virtual-3D (um avatar-3D).

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

1. Com base em requisitos coletados com pessoas surdas, especificar e desenvolver um serviço para tradução automática das legendas (ou do áudio) em língua portuguesa para um vídeo em LIBRAS;
2. Modelar e desenvolver um dicionário de LIBRAS para armazenar uma representação visual de todos os sinais de LIBRAS. Esse dicionário será utilizado pelo serviço de tradução automática na geração do vídeo de LIBRAS;
3. Especificar e desenvolver um serviço para multiplexação do vídeo principal com o vídeo de LIBRAS de forma sincronizada;
4. Validar o protótipo desenvolvido com usuários surdos.

7.3. Detalhamento do Projeto

O serviço proposto é composto por um conjunto de sub-serviços (ou componentes de software) que viabilizarão a geração automática da janela de LIBRAS (isto é, sem a utilização do intérprete), a partir do áudio ou da legenda do vídeo. Na Figura 1 é apresentada uma visão esquemática do serviço proposto.

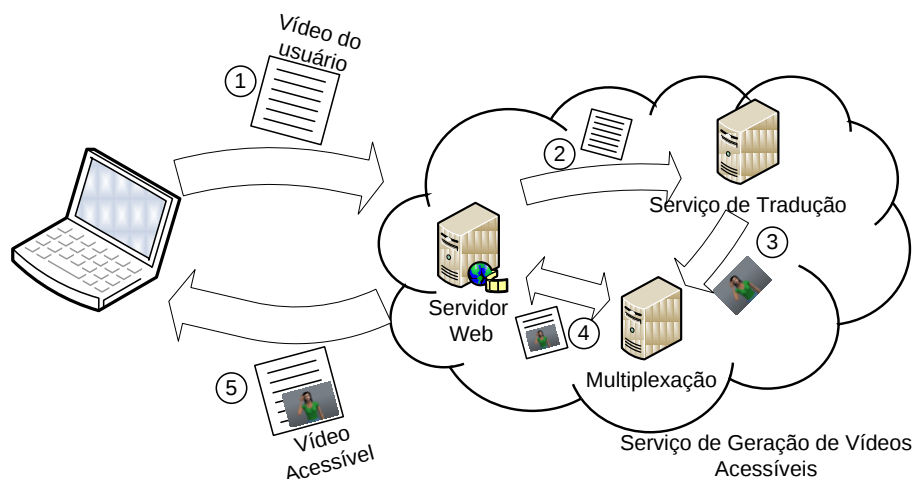


Figura 1. Visão esquemática do serviço proposto

De acordo com a Figura 1, o serviço funcionará da seguinte forma. Inicialmente, um usuário submete um vídeo para o serviço (Figura 1 (1)). Esse vídeo é recebido por um servidor Web que o encaminha para um (sub) serviço de tradução (Figura 1 (2)). O serviço de tradução, como o próprio nome indica, traduz automaticamente as legendas

(ou o áudio) desse vídeo para LIBRAS, produzindo um vídeo (ou legenda) de LIBRAS. O vídeo de LIBRAS é então encaminhado para um (sub) serviço de multiplexação (Figura 1 (3)) responsável por multiplexar o vídeo de LIBRAS e o vídeo do usuário, produzindo um único **vídeo acessível**, isto é, o vídeo do usuário com uma janela (“legenda”) de LIBRAS (Figura 1 (4)). O **vídeo acessível** é então devolvido para o usuário do serviço (Figura 1 (5)).

Além de gerar a janela de LIBRAS de forma automática, o serviço também pode oferecer outros recursos como a possibilidade do usuário configurar a resolução, o tamanho e/ou a posição da janela de LIBRAS dentro do vídeo do usuário, dentre outros. Nas próximas subseções detalharemos os (sub) serviços de tradução e de multiplexação.

7.3.1. Serviço de Tradução

Conforme mencionado anteriormente, o serviço de tradução é responsável por gerar um vídeo de LIBRAS a partir da legenda (ou do áudio) do vídeo em língua portuguesa.

Uma importante característica desse processo é utilização de **Dicionários de LIBRAS** para armazenar representações visuais dos sinais em LIBRAS. Num **Dicionário de LIBRAS**, cada sinal é representado por uma animação ou um vídeo e possui um código associado com essa representação. Dessa forma, o dicionário pode ser definido como um conjunto de tuplas t no seguinte formato:

$$t = \langle g, v \rangle,$$

Onde:

- g é a glosa (ou código) do sinal (i.e., a representação textual do sinal);
- v é a representação visual do sinal;

Como cada sinal possui um código fixo (a *glosa*), a representação visual dos sinais pode ser customizada. Por exemplo, a representação visual pode ser uma animação com um agente animado virtual 3D (um avatar-3D) ou um vídeo gravado por um intérprete. O avatar-3D também pode ser customizado como, por exemplo, um personagem infantil para crianças, etc.

O serviço de tradução também define uma estratégia para sincronizar a janela de LIBRAS com a legenda ou o áudio de entrada. Uma visão esquemática desse serviço é ilustrada na Figura 2.

De acordo com a Figura 2, esse serviço funciona da seguinte forma. Inicialmente, uma filtragem é aplicada vídeo do usuário para extrair os fluxos (mídias) de legenda (ou de áudio) em língua portuguesa (**Filtragem**). Um processo de extração das legendas (ou de reconhecimento de voz) é então aplicado para converter esse fluxo de legendas (ou o áudio) em uma sequência de palavras em formato texto (**Reconhecimento de Voz e Extração de Legendas**). Em seguida, essa sequência de palavras é automaticamente traduzida (**Tradução Automática**) para uma sequência de *glosas* (representação textual de LIBRAS).

Finalmente, um componente de **Exibição** associa cada glosa com sua representação visual armazenada no Dicionário de LIBRAS. Dessa forma, a sequência de *glosas* é convertida para uma sequência de representações visuais que serão sincronizadas para gerar o vídeo de LIBRAS.

A sincronização entre a entrada (legenda ou áudio em língua portuguesa) e a saída (vídeo de LIBRAS) é realizada utilizando o modelo de sincronização baseado no eixo do tempo [8]. Esse modelo define pontos de sincronização que são inseridos no fluxo usando marcações de tempo (*timestamps*) baseadas em um relógio global. Nesse caso, o relógio global é o relógio do arquivo de vídeo. Dessa forma, é possível inserir marcações de tempo no vídeo de LIBRAS de forma sincronizada com as marcações das legendas (ou do áudio).

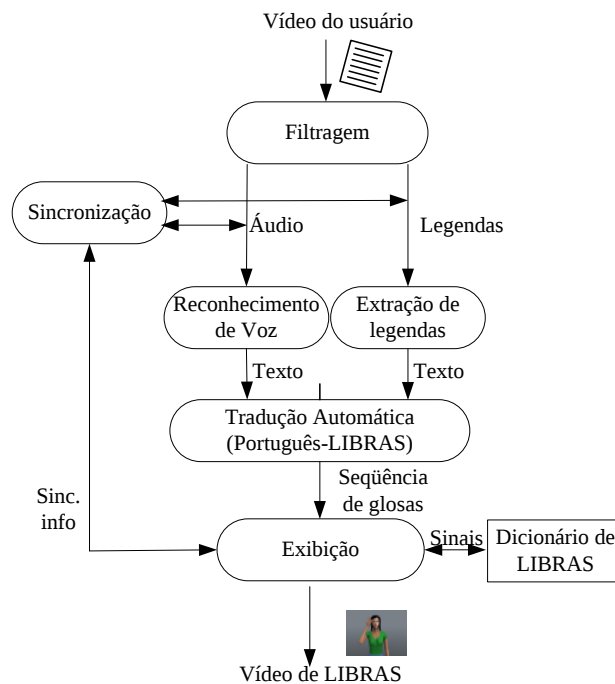


Figura 2. Visão Esquemática do Serviço de Tradução

7.3.2. Serviço de Multiplexação

O serviço de multiplexação, como o próprio nome indica, é responsável por multiplexar o vídeo de LIBRAS gerado pelo serviço de tradução com o vídeo do usuário. Na Figura 3 é apresentada a visão esquemática desse serviço.

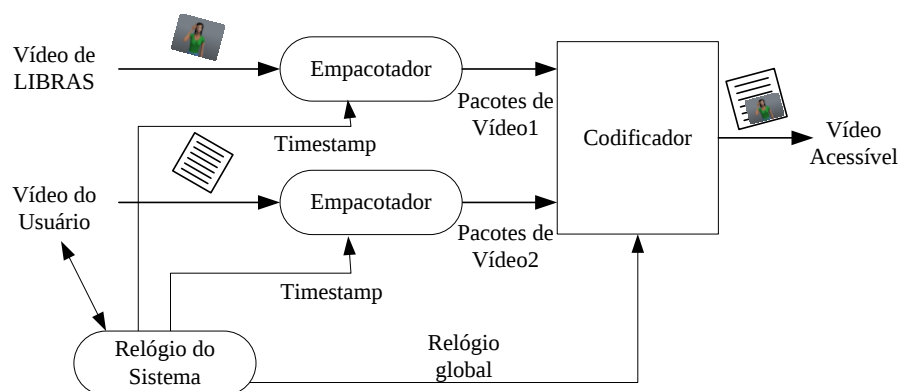


Figura 3. Visão Esquemática do Serviço de Multiplexação

De acordo com a Figura 3, os fluxos de vídeo do usuário e de LIBRAS são, inicialmente, divididos em pacotes para facilitar o processo de multiplexação. Em

seguida, marcações de tempo (*timestamps*) são adicionadas (ou ajustadas) nesses pacotes de acordo com um relógio global compartilhado, i.e., o relógio do vídeo do usuário. Esse **relógio global** é também adicionado no fluxo multiplexado para permitir a decodificação e apresentação sincronizada desses dois fluxos de vídeo. Por fim, um codificador mistura os pacotes dos dois fluxos sincronizados e adiciona periodicamente informações do **relógio global**, gerando um único fluxo de vídeo na saída, o **vídeo acessível**.

8. Ambiente para testes do protótipo

Para validar o serviço proposto, é necessário testá-lo em duas frentes. A primeira frente envolve a validação do serviço, mais especificamente, a validação da inteligibilidade e da usabilidade do sistema, por usuários surdos. A proposta é que a realização desses testes seja realizada utilizando a infra-estrutura do laboratório de usabilidade do CPqD. A segunda frente de validação do serviço envolve um conjunto de testes do serviço sobre uma infraestrutura de rede. Para isso, sugerimos a elaboração de dois protótipos. A seguir, apresentamos a proposta de infra-estrutura para cada um desses protótipos.

Piloto 1: Geração dos conteúdos acessíveis em baixa escala

Cenário: Utilização de um servidor de alto desempenho para rodar o serviço.

Detalhamento: Desenvolver um protótipo do serviço e montar uma infraestrutura para executá-lo em um servidor de alto desempenho no Lavid/UFPB. A infraestrutura proposta prevê o uso da Rede Ipê da RNP para o tráfego dos vídeos (download e upload) e a execução do serviço em um servidor de alto desempenho no Lavid/UFPB. Na Figura 4 é apresentado o ambiente para implantação desse protótipo.

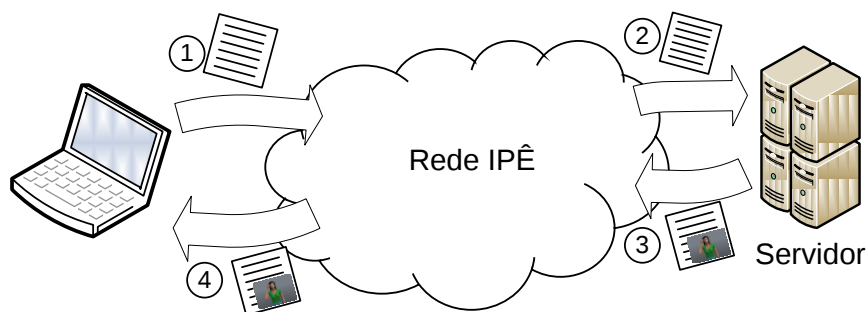


Figura 4. Visão esquemática da infraestrutura do protótipo 1.

Piloto 2: Geração dos conteúdos acessíveis em larga escala

Cenário: Utilização de um serviço de *cloud computing* para rodar o serviço

Detalhamento: Desenvolver um protótipo do serviço e montar uma infraestrutura para executá-lo em um provedor de *cloud computing*. A utilização de um serviço de cloud computing permitiria utilizar os recursos de hardware de acordo com a demanda, aumentando a elasticidade no uso dos recursos. A infraestrutura proposta prevê o uso da Rede Ipê da RNP para o tráfego dos vídeos (download e upload) e o uso de um serviço de cloud computing para executar o serviço. Na Figura 5 é apresentado o ambiente para implantação desse protótipo.

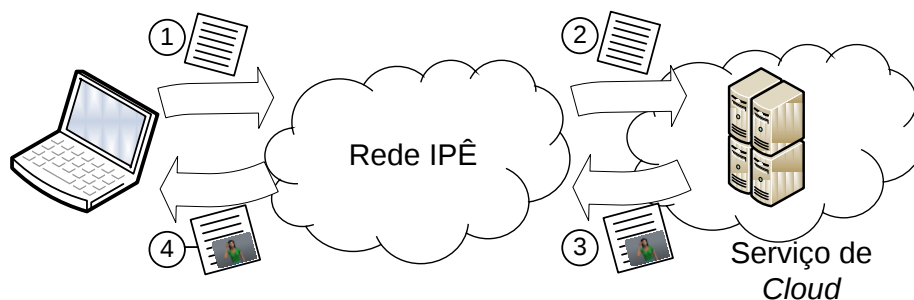


Figura 5. Visão esquemática da infraestrutura do protótipo 2.

9. Referências

- [1] GOES, M. C. R.. **Linguagem, Surdez e Educação**. Campinas: Editora Autores Associados, 1996.
- [2] CAMPOS, M. B.; GIRAFÁ, L. M. M. **SIGNSIM**: uma ferramenta para auxílio à aprendizagem da língua brasileira de sinais. V Congresso Ibero-Americano de informática na Educação – RIBIE. Viñadelmar, Chile: [s.n.]. 2000. p. 1-13.
- [3] STUMPF, M. R.. **Língua de Sinais**: escrita dos surdos na Internet. V Congresso Ibero-Americano de Informática na Educação – RIBIE. Viñadelmar, Chile: [s.n.]. 2000. p. 1-8.
- [4] LEE, D. G.; FELLS, D. I.; UDO, J. P. Emotive captioning. **Computers in Entertainment**, 2007. 3-15.
- [5] LEE, S. E. A. **A gesture based american sign language game for deaf children**. Procceding of Human factors in computing systems - CHI'05. [S.l.]: [s.n.]. 2005. p. 1589-1592.
- [6] STARNER, T.; PENTLAND, A.; WEAVER, J.. Real-Time american sign language recognition using desk and wearable computer based video. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 1998. 1371-1375.
- [7] PICCOLO, L. et al. Modelo de interação inclusivo para interfaces de governo eletrônico, 2010. Fundação CPqD. 87 pp.
- [8] BLAKOWISKI, G.; STEINMETZ, R.. A Media Synchronization Survey: Reference Model, Specification and Case Studies. **IEEE Journal on Selected Areas in Communication**, 14, 1996. 5-35.