Facilita: Auxílio à Leitura de Textos Disponíveis na Web

Willian Massami Watanabe, Renata Pontin de Mattos Fortes, Thiago Alexandre Salgueiro Pardo e Sandra Maria Aluísio
Instituto de Ciências Matemáticas e da Computação
Universidade de São Paulo
Avenida Trabalhador São Carlense, 400
CEP: 13560-970
São Carlos, SP, Brazil
{watinha, renata, taspardo, sandra}@icmc.usp.br

ABSTRACT
Although there are recently multiple ways of delivering media on the web, most web contents are still mainly presented in the textual form. The heavy use of texts pose an accessibility barrier to those who do not read fluently because of both the length and linguistic complexity of the written texts. To deal with that barrier, this paper presents the Facilita, an assistive technology to help lower literacy users to comprehend web applications' textual content. Facilita automatically generates accessible content from web pages, using syntactic simplification and summarization techniques. It is also described the use of technologies aiming at keeping users interactions simple, since lower literacy users, frequently, also show lower computer skills.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: User-centered design; H.5.4 [**H.Hypertext/Hypermedia**]: User issues

General Terms
Design, Human Factors, Languages

Keywords
Acessibilidade na web, Usabilidade, Simplificação Textual

1. INTRODUÇÃO
No Brasil, ainda persistem problemas associados à eficácia escolar, dentre eles, a evasão, a repetência, e a qualidade média da educação ministrada nas escolas brasileiras [6]. O IBGE (Instituto Brasileiro de Geografia e Estatística) relata que, em 2007, 10% dos indivíduos do país com 15 anos de idade ou mais eram analfabetos [6]. Quando este conceito é estendido ao índice de analfabetos funcionais, definido pela UNESCO como o percentual da população com 15 anos de idade ou mais com menos de 4 anos de estudo, o valor atingido de 21,7% da população brasileira.

É importante observar que os analfabetos funcionais possuem dificuldades com atividades que envolvam a leitura, compreensão de textos e extração de informação importante de textos, dependendo do tamanho e da complexidade linguística deles. Logo, possuem acesso restrito aos meios de comunicação que utilizem textos como principal forma de disponibilização de conteúdo.

Nesse contexto, o projeto PorSimples (Simplificação Textual do Português para Inclusão e Acessibilidade Digital[1])[1], considerando esses requisitos de acessibilidade das aplicações, visa o desenvolvimento de tecnologias que facilitem o acesso à informação por indivíduos analfabetos funcionais, por meio de recursos de simplificação textual, sumarização textual e explicitação das informações mencionadas no texto (Reconhecimento de entidades nomeadas, Identificação de funções semânticas e Apresentação das relações retóricas). Esses recursos são aplicados de forma **automática** sobre textos, adaptando-os em textos escritos em linguagem simples, mais facilmente compreendida por indivíduos com dificuldade de leitura.

Este artigo, inserido nos objetivos do projeto PorSimples, tem como finalidade apresentar as características, o desenvolvimento e *design* de interface do **Facilita**, uma aplicação facilitadora de leitura de conteúdos textuais em português de páginas e aplicações Web. Essa aplicação atua como uma tecnologia assistiva[2] que possibilita, ao usuário, a leitura de conteúdos textuais disponíveis na Web. Será enfatizado aqui as operações de simplificação e sumarização textual. A aplicação dessas duas operações sobre o texto gera um conteúdo também textual, com a possibilidade de ser mais facilmente compreendido por usuários analfabetos funcionais.

Este trabalho está organizado da seguinte forma: a próxima seção (Seção 2) descreve as recomendações já abordadas nas diretrizes da WCAG para usuários com dificuldade de leitura; a Seção 3 apresenta a aplicação Facilita, propriamente, juntamente com os seus requisitos de desenvolvimento; por fim, na Seção 4, são apresentadas algumas considerações finais relativas ao que foi realizado no projeto do Facilita, bem como os próximos trabalhos futuros.

[1] http://caravelas.icmc.usp.br/wiki/index.php/Principal
[2] Conjunto de equipamentos, tecnologias ou práticas concebidas e aplicadas para atenuar os problemas encontrados por pessoas com necessidades especiais [3]

2. ABORDAGEM DA WCAG

A WCAG (*Web Content Accessibility Guidelines*), um dos trabalhos mais conhecidos e utilizados no desenvolvimento de aplicações web acessíveis, apresenta recomendações direcionadas a usuários analfabetos funcionais.

A WCAG 2.0, que atingiu o estado de recomendação da W3C em dezembro de 2008, contempla, em seu princípio de compreensibilidade, um requisito de acessibilidade relativo ao nível de escrita das aplicações [12]. Neste requisito (*critério de sucesso* 3.1.5) é especificado que textos que requerem habilidade de leitura superior a de indivíduos com ensino secundário inferior (*lower secondary education*, fazendo menção ao sistema de ensino inglês - quinta a oitava séries no Brasil), devem apresentar uma versão alternativa do mesmo conteúdo ou um conteúdo complementar que seja destinado a indivíduos sem essa habilidade. Por exemplo: disponibilização de sumário textual com conteúdo simplificado; ilustrações, fotos e símbolos para facilitar o entendimento; versões narradas do texto; entre outros.

É importante destacar, no entanto, que essas recomendações de acessibilidade (*Critério de sucesso* 3.1.5 da WCAG 2.0) descreve características relacionadas ao conteúdo da aplicação web. E, para satisfazê-las, torna-se necessária a implementação **manual** desses requisitos na fase de elaboração de conteúdo das aplicações. Ou seja, para garantir a acesso à informação por indivíduos com dificuldade de leitura, os autores do conteúdo devem considerar a utilização desses recursos (utilização de conteúdo simplificado, ilustrações, fotos, símbolos, versões narradas, entre outros) na aplicação. No entanto, para grande maioria dos casos, os responsáveis pela elaboração do conteúdo não são também os responsáveis pelo desenvolvimento do site ou aplicação web. Autores de conteúdo contratam outros indivíduos que sejam desenvolvedores web para implementar seus sites, ou até mesmo, no contexto da Web 2.0, os próprios usuários são disseminadores de informação nas aplicações web e por isso, muitas vezes, não possuem conhecimento a respeito das diretrizes de acessibilidade na web [5].

Um outro aspecto a ser destacado diz respeito ao fato de que usuários com baixo nível de alfabetização, são, muitas vezes também, classificados como analfabetos computacionais, portanto requerem interações simples de utilização, dadas sua dificuldade e pouca familiaridade com o uso de sistemas computacionais. A simples disponibilização de recursos como ilustrações, fotos, símbolos ou versões narradas do texto pode não garantir a acessibilidade da aplicação à usuários classificados como analfabetos funcionais. Faz-se necessária a implementação de uma interface com interações de uso simplificadas, adequadas ao nível de conhecimento computacional agregado por esses indivíduos, para permitir e facilitar o uso dessa aplicação.

3. O PROJETO FACILITA

O "Facilita" é uma aplicação web que funciona como uma tecnologia assistiva de auxílio à leitura de conteúdos textuais disponíveis em páginas web. O desenvolvimento dessa aplicação está inserido no projeto PorSimples, e deve utilizar principalmente recursos de simplificação e sumarização automática de textos para promover o acesso à informação por indivíduos analfabetos funcionais.

Esse auxílio à leitura é realizado por meio de duas adaptações **automáticas** [2] de textos disponíveis na web: a simplificação sintática que gera textos mais facilmente lidos e compreendidos [7] e a sumarização textual (redução no número de sentenças) que gera textos mais curtos e mais facilmente lidos pelo público alvo do projeto [9]. Ao resultado dessas operações é atribuído o nome de **Conteúdo Facilitado**.

É importante destacar os requisitos de *design* da interface, considerando que o público alvo dessa aplicação (analfabetos funcionais) são muitas vezes também analfabetos computacionais. Dessa forma, é necessário o desenvolvimento de uma interface levando em conta as restrições desses usuários, implementando recomendações de acessibilidade e também princípios de *design* para usuários com baixo nível de alfabetização [13], para possibilitar o acesso às funcionalidades da aplicação.

3.1 O Conteúdo Facilitado

Conforme mencionado, o Facilita é uma aplicação que realiza a adaptação automática de conteúdos textuais disponíveis em sites e aplicações web em Conteúdo Facilitado, acessível a usuários analfabetos funcionais. Esse Conteúdo Facilitado é um conteúdo também textual, no entanto sua estrutura será simplificada, conforme definido no projeto PorSimples com base na *Plain Language*[3], sistemas de simplificação para o Inglês [11] e na análise de corpus de textos reescritos para atender um público mais amplo na Web Brasileira [1]. O tamanho do texto também será reduzido, visto a dificuldade de usuários com baixo nível de alfabetização com textos extensos.

Para realizar a elaboração desse conteúdo de forma **automática**, são utilizadas as operações de "Simplificação" e "Sumarização" textual, conforme descritas a seguir:

Simplificação: o objetivo da operação de simplificação é simplificar a estrutura linguística dos textos no intuito de maximizar sua compreensão [7]. Qualquer processo que reduza a complexidade léxica ou sintática de um texto, preservando o seu significado e as informações do seu conteúdo, pode ser definido como Simplificação textual. O objetivo da Simplificação textual é facilitar a compreensão do texto por humanos ou o seu processamento por outros programas [11]. Para elaborar textos simplificados, os autores podem remover sentenças ou frases, dividir sentenças extensas em múltiplas sentenças, modificar o vocabulário, encurtar frases longas descritivas, entre outras [10]. Especificamente no Facilita, são utilizadas regras de simplificação definidas no Manual de Simplificação Sintática para o Português, criado no projeto PorSimples [1].

Para exemplificar o funcionamento da operação de simplificação será utilizado o texto a seguir:

> Conforme mapas climatológicos do Centro de Previsão de Tempo e Estudos Climáticos (Cptec), ligado ao Instituto Nacional de Pesquisas Espaciais (Inpe), desde 1995 a média dos picos diários de calor não era tão elevada em dezembro.

[3] http://www.plainlanguage.gov/

Após a aplicação da operação de simplificação textual desenvolvida no projeto, é obtido o seguinte texto:

> Conforme mapas climatológicos do Centro de Previsão de Tempo e Estudos Climáticos (Cptec), desde 1995 a média dos picos diários de calor não era tão elevada em dezembro. O Centro de Previsão de Tempo e Estudos Climáticos (Cptec) é ligado ao Instituto Nacional de Pesquisas Espaciais (Inpe).

Nesse exemplo, foi efetuada a operação de separação do aposto presente no primeiro texto ("ligado ao Instituto Nacional de Pesquisas Espaciais") da oração principal.

Sumarização: o objetivo da sumarização automática é, considerando uma fonte de informações, extrair seu conteúdo, e apresentar os conteúdos mais importantes ao usuário em uma forma condensada e sensível às necessidades do usuário ou da aplicação [8]. No Facilita é utilizado o método EPC-P (Extração de Palavras-Chave por Padrão), que teve melhor precisão na avaliação realizado por Margarido et al. [9]. Esse método avalia a presença de palavras-chave nas sentenças do texto. As sentenças que possuírem palavras-chave são mantidas no sumário.

A ordem de execução das operações de processamento textual foi definida pelos integrantes do projeto PorSimples como sumarização seguida de simplificação.

Além dos analfabetos funcionais, outros grupos de usuários podem ser beneficiados por essas operações, como por exemplo: usuários com afasia, disléxicos, aprendizes de uma língua estrangeira, idosos, surdos e crianças e adultos em fase de letramento.

A próxima seção (Seção 3.2) descreve as características do protótipo do Facilita.

3.2 Protótipo do Facilita

Foi desenvolvido um protótipo do Facilita durante o projeto, para que fossem realizadas avaliações sobre a forma de interação com o usuário que deve ser utilizada e também sobre os aspectos de *design* da interface. Esse protótipo implementa a operação de Simplificação e a Sumarização com 70% de compressão do texto.

Para que o Facilita realize esse auxílio à leitura, devem ser realizados os seguintes passos, durante uma interação do usuário com a aplicação:

1. O usuário deve acessar um site ou determinada aplicação web.
2. O usuário deve selecionar o conteúdo textual disponível no site que se deseja auxílio na leitura.
3. O usuário deve clicar no *link* do Facilita disponível na barra de favoritos do navegador (ilustrado na Figura 1) e, em seguida, lhe será apresentado o conteúdo **Facilitado** (ilustrado na Figura 2).

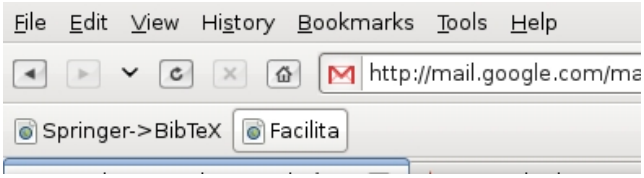

Figura 1: Barra de favoritos do navegador com o *link* do Facilita

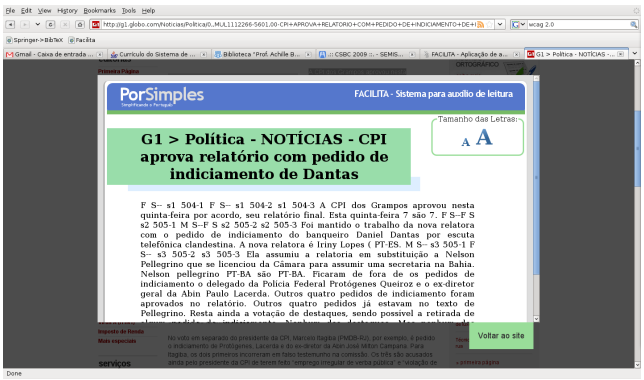

Figura 2: Apresentação do conteúdo Facilitado inserido no elemento popup.

Essa abordagem para interação com o usuário é bastante semelhante à utilizada pelo sistema Translate da Google[4], em que o usuário também deve selecionar textos disponíveis em sites e clicar em um *link* instalado na barra de favoritos no seu navegador. No entanto, o sistema Translate da Google redireciona o site que estava sendo acessado para a aplicação Translate, enquanto o Facilita realiza a disponibilização do conteúdo gerado no próprio site, por meio de um componente de interface *popup*.

A inserção da funcionalidade do Facilita como parte do navegador (inserção do *link* do Facilita na barra de favoritos) foi identificada como uma das melhores formas de ativação do recurso, uma vez que sua visualização é constante ao usuário. Ou seja, independentemente do site que esteja sendo acessado, o recurso sempre é apresentado em um mesmo lugar da interface.

Após o procedimento de ativação do recurso, é apresentado um componente de interface *popup* no meio da janela do navegador, como no exemplo de execução do Facilita ilustrado na Figura 2. É importante destacar que a utilização da componente de interface *popup* foi realizada como uma forma de manter informações contextuais relacionadas ao site ou aplicação web sobre o qual foi aplicada a Facilitação do conteúdo. Ao mesmo tempo em que é apresentado o conteúdo Facilitado dentro do elemento *popup*, pode ser observado, ao fundo, o site ou aplicação web que estava sendo acessado anteriormente. Esse componente *popup* foi implementado utilizando elementos *div* do HTML, sem a criação de uma nova janela, visto que o aumento do número de janelas sendo apresentadas pelo sistema operacional poderia atrapalhar o usuário. No canto direito inferior da *popup*,

[4]http://translate.google.com/translate_tools

é disponibilizado um *link* que fecha a *popup*, e retorna ao site que estava sendo acessado.

Foi observado que as operações de Simplificação e Sumarização textual precisa de alguns segundos para ser completada, de acordo com o tamanho do texto que for selecionado nos sites. Durante o tempo de processamento necessário às operações de Simplificação e Sumarização , é apresentada uma mensagem de "Aguarde" ao usuário.

4. CONSIDERAÇÕES FINAIS

Este trabalho discutiu as questões de projeto da interação e o desenvolvimento de uma tecnologia assistiva, destinada a usuários com dificuldade de leitura (analfabetos funcionais) de conteúdos textuais em português que são disponibilizados em sites e aplicações web.

Foi descrita a geração automática de conteúdo acessível por meio de operações de processamento textual (Simplificação e Sumarização textual), juntamente com abordagens no *design* de interfaces para usuários com baixos níveis de alfabetização, que, muitas vezes, possuem também pouca experiência e habilidade na utilização de computadores.

Ainda não foram realizadas avaliações de usabilidade sobre a aplicação desenvolvida neste trabalho. Dessa forma, como trabalhos futuros, devem ser conduzidas metodologias de inspeção (avaliação heurística e percurso cognitivo) e testes de usabilidade, para garantir que os usuários foco do projeto (analfabetos funcionais) sejam capazes de utilizar a aplicação.

No projeto PorSimples, também estão sendo realizados estudos empíricos a respeito da efetividade do Conteúdo Facilitado como forma de disponibilização de informação para os usuários.

Por fim, também está sendo desenvolvida uma versão do Facilita que, além de utilizar as operações de Simplificação e Sumarização textual, utilizará a elaboração textual [14], que consiste da explicação de informações mencionadas no texto. Essa versão do Facilita será destinada para atividades educativas.

5. AGRADECIMENTOS

Os autores agradecem a todos os membros do projeto PorSimples que diretamente ou indiretamente auxiliaram no desenvolvimento deste trabalho e à FAPESP e Microsoft Research (MSR) pelo apoio financeiro.

6. REFERÊNCIAS

[1] S. M. Aluisio, L. Specia, T. A. Pardo, E. G. Maziero, and R. P. Fortes. Towards brazilian portuguese automatic text simplification systems. In *DocEng '08: Proceeding of the eighth ACM symposium on Document engineering*, pages 240–248, New York, NY, USA, 2008. ACM.

[2] J. Burstein. Opportunities for natural language processing research in education. In *CICLing '09: Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing*, pages 6–27, Berlin, Heidelberg, 2009. Springer-Verlag.

[3] A. M. Cook and S. M. Hussey. *Assistive technologies: Principles and practice*, volume 2a ed. Mosby - Year Book, 2001.

[4] A. P. Freire, R. Goularte, and R. P. de Mattos Fortes. Techniques for developing more accessible web applications: a survey towards a process classification. In *SIGDOC '07: Proceedings of the 25th annual ACM international conference on Design of communication*, pages 162–169, New York, NY, USA, 2007. ACM.

[5] A. P. Freire, C. M. Russo, and R. P. de Mattos Fortes. The perception of accessibility in web development by academy, industry and government: a survey of the brazilian scenario. *New Review of Hypermedia and Multimedia*, 14(2):149–175, December 2008.

[6] IBGE. Síntese dos indicadores sociais - uma análise das condições de vida da população brasileira, 2008. Available at: http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/indicadoresminimos/sinteseindicsociais2008/indic_sociais2008.pdf.

[7] A. C. Jr., E. Maziero, C. Gasperin, T. A. S. Pardo, L. Specia, and S. M. Aluisio. Supporting the adaptation of texts for poor literacy readers: a text simplification editor for brazilian portuguese. In *Workshop Innovative Use of NLP for Building Educational Applications*, pages 34–42. North American Association for Computational Linguistics - Human Language Technologies conference, Boulder, Colorado, June 2009.

[8] I. Mani. *Automatic summarization.* John Benjamins Publishing Company, 2001.

[9] P. R. A. Margarido, T. A. S. Pardo, and S. M. Aluísio. Sumarização automática para simplificação de textos: Experimentos e lições aprendidas. In *Anais do Workshop de Usabilidade, Acessibilidade e Inteligibilidade aplicadas em interfaces para analfabetos, idosos e pessoas com deficiência, 2008*, pages 1–6. In: Workshop de Usabilidade, Acessibilidade e Inteligibilidade aplicadas em interfaces para analfabetos, idosos e pessoas com deficiência, 2008, Porto Alegre, 2008.

[10] S. E. Petersen. Syntactic simplification. Phd Thesis - University of Washington, 2007.

[11] A. Siddharthan. Syntactic simplification and text cohesion. *Research on Language and Computation*, 4(1):77–109, June 2006.

[12] W3C. Web content accessibility guidelines (wcag) 2.0. W3C Recommendation, December 2008. Available at: http://www.w3.org/TR/WCAG20/

[13] W. M. Watanabe and R. P. de Mattos Fortes. Revisão sistemática sobre princípios de design de aplicações web acessíveis para analfabetos funcionais. In the Proceedings of *SEMISH - XXXVI Seminário Integrado de Software e Hardware, 2009*. CSBC - XXIX Congresso da Sociedade Brasileira da Computação, pages 403-417, 2009.

[14] D. N. Young. Linguistic simplification of SL reading material: effective instructional practice? *The Modern Language Journal*, 83(3):350–366, 1999.