

# Conversão de texto para hipertexto: um processo para retextualização digital

Antônio Robson de Paula , Ivan L. M. Ricarte (Orientador)

Departamento de Engenharia de Computação e Automação Industrial (DCA)

Faculdade de Engenharia Elétrica e de Computação (FEEC)

Universidade Estadual de Campinas (Unicamp)

Caixa Postal 6101, 13083-970 — Campinas, SP, Brasil

{depaula, ricarte}@dca.fee.unicamp.br

**Abstract** – This paper presents an iterative process to convert linear text to hypertext, an activity named digital retextualization. State of the art hypertext authoring systems were evaluated as potential digital retextualization tools, but they lack the support to important steps required to this process. Thus, an open source tool to support text to hypertext conversion was developed. In this way, it is possible to validate the proposed digital retextualization process while allowing other researchers to continue this work.

**Keywords** – Hypertext, text, digital media, links, retextualization.

## 1. Introdução

O formato de uma mensagem depende do meio utilizado para a sua transmissão. O conteúdo de uma apresentação oral não pode ter exatamente a mesma forma que sua apresentação impressa em papel. Assim também o meio digital demanda suas próprias convenções para a apresentação de conteúdos. No entanto, escrever hipertextos não é uma tarefa natural para autores habituados a escrever textos tradicionais, com estrutura linear e idéias organizadas hierarquicamente.

No texto tradicional, a ordem de leitura é estabelecida pela mídia impressa e, mesmo que o leitor adote uma seqüência de leitura não linear, ele o faz conhecendo o contexto originalmente proposto pelo autor. Hipertextos, por outro lado, são fragmentados e têm estrutura não linear [3]. Os segmentos de texto conectados por hiperligações precisam ser escritos de forma a preservar a coerência, mesmo que a leitura ocorra em uma ordem diferente da proposta pelo autor. Assim, um procedimento para escrita de hipertextos que considere tais aspectos pode auxiliar seus autores.

O hipertexto estabelece uma nova topografia de escrita e, como outras tantas que foram introduzidas ao longo da história humana, leva a novos processos de autoria, revelados após sucessivas traduções de uma topografia original para uma nova topografia de escrita [2]. Se desenvolvermos um modelo de conversão de texto para hipertexto, este poderá ser utilizado para posteriormente definir um procedimento formal para a escrita de hipertextos.

A experiência existente na lingüística para

a conversão de textos orais para textos escritos, um processo denominado retextualização [8], pode ser aproveitada para entender o processo de conversão de texto linear para hipertexto, aqui denominado *retextualização digital*. O estudo da teoria envolvida no processo de retextualização auxiliou na identificação de operadores para a retextualização digital, permitindo propor um modelo para a transformação do texto linear em hipertexto. Com tal modelo, foi possível avaliar os aplicativos para autoria de hipertexto existentes, o que motivou o desenvolvimento de uma ferramenta de código aberto para apoiar a retextualização digital.

## 2. Retextualização digital

A conversão de um texto oral para escrito é precedida por um processo de transcrição ou transcodificação [7], que é a passagem do material fonético para o meio impresso utilizando operações simples como a eliminação da hesitação [11]. A retextualização pode ser entendida como a interferência no código e no sentido do texto transcrito de maneira a preservar a informação na nova topografia [8].

Emprestando estes conceitos da lingüística, podemos definir que a retextualização digital também deve ser precedida de uma transcrição digital, que é o trabalho de converter, para o meio digital, um texto linear preparado para o meio impresso. Tal atividade é facilitada pela ampla adoção do computador como ferramenta para autoria de textos.

A retextualização digital é um processo ainda indefinido, porém acreditamos que conduzindo várias conversões de texto para hipertexto po-

deremos identificar operações que auxiliarão na sua definição. O exemplo a seguir [4] ilustra o tipo de conversão que experimentamos. Considere o seguinte texto extraído de uma revista semanal: “Um dos maiores investigadores do poder da mente sobre a saúde é o cardiologista americano Herbert Benson, da Universidade Harvard, autor do livro Medicina Espiritual. Pesquisas conduzidas por ele mostram que, em média, 60% das consultas médicas poderiam ser evitadas, caso as pessoas usassem sua capacidade mental para combater naturalmente tensões que são causadoras de problemas físicos.” Para converter este texto para hipertexto, os autores propõem as operações de segmentação das informações textuais, de introdução de redundância para tornar fragmentos coerentes, e de definição de associações entre fragmentos. O resultado obtido foi o seguinte:

**Herbert Benson:** um dos maiores investigadores do poder da mente sobre a saúde é o cardiologista americano Herbert Benson, da Universidade Harvard, autor do livro Medicina Espiritual.

**Poder da mente:** Pesquisas conduzidas por Benson mostram que, em média, 60% das consultas médicas poderiam ser evitadas, caso as pessoas usassem sua capacidade mental para combater naturalmente tensões que são causadoras de problemas físicos.

Observe que no fragmento de título “Poder da mente”, o pronome *ele* foi modificado para *Benson* de maneira a preservar o entendimento do fragmento fora do contexto original, além de ter sido definido como uma âncora que dispara a ligação para o fragmento de título “Herbert Benson”. Este exemplo mostra que interferir no código e no significado é necessário para converter um texto linear em hipertexto.

Ao realizar algumas retextualizações digitais, notamos a necessidade dos seguintes passos: fragmentação, aglutinação, definição de perspectivas do texto, associação da nomenclatura da perspectiva ao fragmento (classificação), interferência no conteúdo do fragmento, criação de ligações de conteúdo, criação de propostas de seqüências de fragmentos (caminhos de leitura), com definição de fragmentos para início das leituras (raízes). Assumimos que estes passos, executados iterativamente, estabelecem um modelo para a retextualização digital. A Fig. 1 ilustra tal processo.

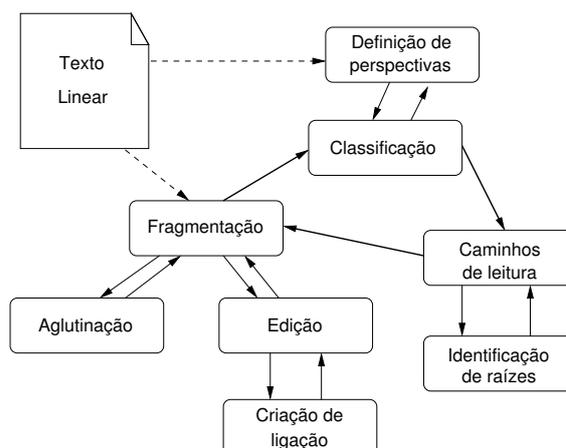


Figura 1. Processo de retextualização digital.

Na próxima seção, sistemas existentes para autoria de hipertexto são avaliados como potenciais ferramentas para realizar a retextualização digital.

### 3. Avaliação de sistemas existentes

O uso de ferramentas automáticas para conversão de texto em hipertexto não é adequado no contexto da retextualização digital, uma vez que o processo de escrita de hipertexto requer pensamento associativo e tomadas de decisões subjetivas [5]; assim sendo, essa categoria de sistemas não foi avaliada.

No contexto de sistemas para autoria de hipertexto e com o fato de que a Web é hoje um recurso recorrente no dia à dia das pessoas, consideramos que um sistema de retextualização digital deve possuir a funcionalidade de publicação na Web. Adicionalmente, as ligações criadas pelas ferramentas de autoria devem poder disparar uma página Web, fora do texto sendo convertido, através de sua URL (*Uniform Resource Locator*); este tipo de ligação é denominada de ligação externa.

Foram avaliadas cinco ferramentas de autoria de hipertextos. Viki [9] definiu o termo hipertexto espacial referindo-se ao uso de uma representação gráfica dos fragmentos de texto para aprimorar a expressão da estrutura do hipertexto para o autor. HART [10] foi desenvolvida para facilitar o processo de conversão de texto para hipertexto. Microsoft Word é um editor de textos usualmente utilizado para escrever documentos para a Web. Dreamweaver é um pacote de criação de páginas para a Web. Por fim, Storyspace [1] é a ferramenta adotada pela comunidade de hipermídia como padrão *de facto* para experimentos de hipertexto.

Passos da retextualização digital	VIKI	HART	MS Word	Dreamweaver	Storyspace
Fragmentação	✓	✓	✓	✓	✓
Aglutinação	✓	x	✓	✓	✓
Definição das perspectivas do texto	x	x	x	x	x
Classificação	x	x	x	x	x
Interferência no conteúdo do fragmento	✓	x	✓	✓	✓
Criação de ligações de conteúdo	x	Exceto Web	✓	✓	Exceto Web
Caminhos de leitura	x	x	x	x	x
Definição de fragmento raiz	x	x	x	x	x

**Tabela 1. Resultado da avaliação.**

O resultado dessa avaliação está resumido na Tab. 1, onde podemos notar que as ferramentas avaliadas não dão suporte aos passos de definição das perspectivas do texto e classificação que auxiliam o autor no momento da escrita do hipertexto. Nota-se também a falta de suporte pelas ferramentas avaliadas aos passos de criação dos caminhos de leitura propostos pelo autor e definição de suas raízes. Estes passos são essenciais para que se obtenha um hipertexto que possa ter múltiplas leituras coerentes.

Esta avaliação mostrou que os sistemas atuais para autoria de hipertextos não podem ser utilizados no processo iterativo de retextualização digital. Desta forma, o projeto e implementação de uma ferramenta para este fim tornou-se necessário.

#### **4. Ferramenta de retextualização digital**

Nesta seção descreveremos brevemente a ferramenta para retextualização digital de código aberta denominada TEXTHIT (*Text to Hypertext Conversion Tool*), desenvolvida no contexto deste trabalho. O objetivo desta ferramenta é permitir que autores de hipertexto avaliem a retextualização digital e contribuam com o desenvolvimento da ferramenta e do modelo proposto.

##### **4.1. Cenário de uso**

Nesta seção um cenário de uso do TEXTHIT é apresentado, descrevendo o processo de retextualização digital de um texto já transcrito para o meio digital (no formato HTML) e considerando que o resultado será publicado na Web.

Inicialmente, o autor seleciona um arquivo contendo o texto linear. A ferramenta executa uma segmentação automática, produzindo um fragmento

(com um título padrão) para cada parágrafo. O texto original permanece disponível para consulta do autor durante o processo.

Na seqüência, o autor define as perspectivas para o texto, ou seja, quais são os temas abordados no texto que podem motivar uma leitura diferenciada. A cada perspectiva é associada uma cor, que será utilizada para classificar os fragmentos. Os fragmentos obtidos na inicialização do texto são apresentados em cinza, indicando que ainda não foram classificados. Com essas perspectivas o autor classifica cada fragmento, podendo também alterar o título atribuído ao fragmento.

Nem sempre o segmento sob análise contempla uma única perspectiva. Para adequá-lo, o autor pode editar seu conteúdo, re-fragmentando o texto ou aglutinando fragmentos. Ligações, internas e externas, podem ser criadas nessa etapa de edição.

Com os fragmentos classificados e as perspectivas definidas, o autor pode propor os caminhos de leitura que fazem sentido para o seu texto. Cada um desses caminhos tem um ponto para o início da leitura naquela perspectiva, que é a raiz do caminho de leitura e que precisa ser definido pelo autor. Para auxiliá-lo nessa tarefa, a ferramenta oferece uma visualização do caminho de leitura.

Ao avaliar os caminhos de leitura, o autor pode rever suas definições iniciais, alterando fragmentos e perspectivas. Quando satisfeito com o resultado obtido, ele pode publicar seu trabalho na Web, indicando o endereço de um servidor que recebe o conjunto de fragmentos como uma aplicação Web.

## 4.2. Estado da implementação

A ferramenta TEXTHIT foi desenvolvida em Java, sendo organizada em uma arquitetura de três camadas (interface com o autor, motor de retextualização e persistência) e utilizando o padrão de desenvolvimento MVC [6].

A interface com o autor contém os componentes: editor de mapas, para exibir a representação espacial do hipertexto e permitir a classificação de fragmentos; editor de nós, para edição, criação e aglutinação dos fragmentos de texto; visualizador de nós, para pré-visualização do conteúdo dos fragmentos; editor de caminhos de leitura, para criação de possíveis seqüências de leitura; e publicador de mapas, para definição dos dados relativos à publicação do hipertexto em um servidor Web.

Na camada motor de retextualização ocorre o processamento das requisições da camada de interface, mantendo as informações relativas ao modelo de hipertexto definido. Na última camada há um componente para a persistência de nós, para a reutilização de um fragmento em mais de um hipertexto; e outro para a persistência de mapas, responsável por permitir a publicação e a retomada do trabalho em sessões futuras.

O código do TEXTHIT é livre para ser modificado pela comunidade e está disponível na página do projeto<sup>1</sup>.

## 5. Conclusão

Este trabalho apresentou um modelo para retextualização digital que tem por objetivo iniciar o processo de formalização da autoria de hipertextos. A fim de apoiar esse processo, foi desenvolvida uma ferramenta para validar e futuramente refinar o modelo proposto. Para tanto, a ferramenta tem sido disponibilizada a parceiros da área de lingüística que trabalham com textos no meio digital.

A ferramenta teve de ser desenvolvida porque os principais sistemas para autoria de hipertexto não são de código aberto e portanto não permitem alterações; a opção por disponibilizar esta ferramenta com código aberto procura evitar que futuros desenvolvimentos nesta área precisem repetir esse esforço.

## Referências

- [1] M. Bernstein. Storyspace 1. In *Proc. 13th ACM Conference on Hypertext and Hypermedia*, pages 172–181, New York, NY, USA, 2002. ACM Press.
- [2] J. D. Bolter. Topographic writing: Hypertext and the electronic writing space. In *Hypermedia and literary studies*, pages 105–118, Cambridge, MA, USA, 1991. MIT Press.
- [3] D. B. Braga. Hipertexto: questões de produção e leitura. In *Estudos Linguísticos*, volume XXXIV, pages 756–761, 2005.
- [4] D. B. Braga and I. L. M. Ricarte. Letramento na era digital: construindo sentidos através da interação com hipertextos. In *Revista da Anpoll*, volume 18, pages 59–82, 2005.
- [5] C. H. Franke III and N. J. Wahl. Authoring a hypertext UNIX help manual. In *Proc. ACM 23rd Annual Conference on Computer Science*, pages 238–245, New York, NY, USA, 1995. ACM.
- [6] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns — Elements of Reusable Object-Oriented Software*. Addison Wesley Longman, Reading, MA, USA, 1998.
- [7] J. J. Gumperz and N. Berenz. Transcribing conversational exchanges. Technical Report COGSCI-90-63, 1990.
- [8] L. A. Marchuschi. *Da fala para a escrita: atividades de retextualização*. Cortez, São Paulo, SP, Brasil, 2001.
- [9] C. C. Marshall, F. M. Shipman III, and J. H. Coombs. Viki: Spatial hypertext supporting emergent structure. In *Proc. ACM European Conference on Hypermedia Technology*, pages 13–23, New York, NY, USA, 1994. ACM.
- [10] J. Robertson, E. Merkus, and A. Ginige. The hypermedia authoring research toolkit (hart). In *Proc. ACM European Conference on Hypermedia Technology*, pages 177–185, New York, NY, USA, 1994. ACM.
- [11] R. Rojo and B. Schneuwly. As relações oral/escrita nos gêneros orais formais e públicos: o caso da conferência acadêmica. In *Revista Linguagem em (Dis)curso*, volume 6, pages 463–493, Tubarão, SC, Brasil, 2006.

<sup>1</sup><http://code.google.com/p/texthit/>