

# Avaliação de hiperdocumentos a partir da classificação de *links* \*

*Autor(a): Elisandra Aparecida Alves da Silva*

*Orientadora: Renata Pontin de Mattos Fortes*

{elisilva, renata}@icmc.sc.usp.br

Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo  
Caixa Postal 668, São Carlos - SP, CEP 13560-970

## Resumo

A criação de *links* é uma atividade essencial durante a autoria de um hiperdocumento. Os *links* são um diferencial importante em relação à autoria convencional de documentos e por meio deles, é possível a interação básica, conhecida como "navegação" pelo hiperdocumento com o suporte de um *browser*. Durante a construção de um hiperdocumento, os autores são projetistas dessa estrutura bastante flexível de *links* que, em geral, fica propensa a "erros" por envolver um grande número de páginas com enorme diversidade de tipos de *links*, e suas respectivas formas de interação. A proposta deste trabalho de mestrado é especificar uma ferramenta que possa auxiliar a avaliação de hiperdocumentos utilizando um método eficiente de classificação de *links*.

## Abstract

The creation of links is an essential activity during the hypertext construction. The links are an important differential in relation to conventional construction of documents and, through them, the basic interaction is possible, known as "navigation" through the hypertext using a browser as a support. During the hypertext construction, the authors are designers of this flexible structure of links. In general, this structure is inclined to "errors" because it involves a great number of pages with an enormous diversity of kinds of links, and their respective ways of interaction. The purpose of this master essay is to specify a tool which could help the evaluation of hypertexts using an efficient method for links classification.

**Palavras-chave:** classificação de *links*, hiperdocumento, qualidade.

---

\* Projeto desenvolvido com o apoio financeiro da FAPESP

## 1. Introdução

O projeto de um hiperdocumento não é um processo exato, e são muitos os diferentes aspectos envolvidos no problema de produção de um “bom” hiperdocumento que devem ser considerados, de acordo com (Fortes & Nicoletti, 1997). O autor quando projeta um hiperdocumento, por sua vez, se depara com as atividades de projeto de uma aplicação hipermídia, considerada como uma atividade de desenvolvimento de um software. Mas diferentemente do projeto de um software, ao projetar o hiperdocumento, o autor produz maneiras alternativas de organizar as informações, pois não existem regras consolidadas a serem seguidas para a execução de um bom projeto. A inexistência de procedimentos de avaliação sistemáticos, que possam ser utilizados durante a criação de um hiperdocumento, torna a tarefa de autoria “desamparada”, uma vez que quais os aspectos que demonstram qualidade, em geral, somente podem ser observados após intensa interação por parte dos usuários. Essa pode ser uma razão para a existência de hiperdocumentos constatados como “mal escritos”.

Com relação ao software em geral, (Ebert, 1996) afirma que “metas como melhoria da qualidade em termos de confiabilidade e manutenibilidade são de grande importância no desenvolvimento de software”. Nesse sentido, essas propriedades também podem ser almejadas no processo de autoria de hiperdocumentos. Esses objetivos podem ser alcançados, durante o processo de autoria, pelo uso de uma ferramenta que forneça *feedback* sobre a estrutura do hiperdocumento.

Nesse contexto, pretendemos auxiliar a avaliação de hiperdocumentos da *World-Wide-Web* por meio de uma ferramenta de suporte à classificação de seus *links*. Para esta classificação alguns sistemas de aprendizado de máquina da área de Inteligência Artificial deverão ser utilizados.

Na Seção 2, algumas métricas de qualidade de hiperdocumentos são apresentadas, e na Seção 3 a abordagem de avaliação de *links* da *Web* que pretendemos investigar. As considerações finais encontram-se na Seção 4.

## 2. Métricas de qualidade de hiperdocumentos

Um dos primeiros trabalhos a enfatizar a importância de se avaliar a qualidade de hiperdocumentos foi realizado por (Brown, 1990). Ele concluiu, numa primeira instância, que a avaliação deve sempre contar com o discernimento e bom senso das pessoas envolvidas no processo de desenvolvimento, mas que o auxílio de ferramentas automáticas que analisem as características inerentes ao hiperdocumento é um suporte fundamental para essa atividade humana. (Blustein et al., 1997) também considera que todas as abordagens para avaliar a qualidade de *links* envolvem diretamente avaliação humana ou mecânica.

De acordo com (Hatzimanikatis et al, 1995), não existe uma definição completa de qualidade de hiperdocumento. As metodologias e ferramentas que poderiam ajudar na análise e avaliação da qualidade de hiperdocumentos ainda não fornecem o auxílio que deveriam. A possibilidade de medir atributos de um hiperdocumento é imprescindível para as atividades de autoria e Engenharia de Hipertexto.

O uso de métricas baseadas em *links* das quais valores pudessem ser coletados periodicamente por uma ferramenta automática pode contribuir para uma boa administração do processo de autoria de hiperdocumentos e manutenção da *Web*. Métricas baseadas em *links* podem ser usadas para identificar algumas categorias de *links* que podem, eventualmente, causar problemas na qualidade.

As primeiras métricas de hiperdocumentos em geral foram propostas por (Botafogo & Shneiderman, 1992), e posteriormente usadas por (Rivlin et al., 1994). Elas se baseiam

prioritariamente na aplicação de técnicas baseadas em grafos para obtenção das medidas relativas a: a) **compactação**, que indica a conectividade intrínseca do hiperdocumento. Do ponto de vista do leitor, um alto valor de compactação indica que cada nó possui muitos *links*; b) **estratificação**, que revela o grau de organização de um hiperdocumento; o máximo valor de estratificação ocorre quando o hiperdocumento é linear.

Os fatores de qualidade importantes, identificados como legibilidade e manutenibilidade de um hiperdocumento, bem como os critérios correspondentes foram definidos por Hatzimanikatis. Entretanto, estes dois fatores não podem ser medidas obtidas diretamente. Por isso, esses fatores são decompostos em critérios de nível mais baixo. A maioria dos critérios afeta ambos os fatores, a saber: **tamanho, complexidade do caminho, impureza da árvore, modularidade, complexidade individual do nó, coerência, complexidade do conteúdo dos nós, simplicidade**.

Um outro método de avaliação, conhecido como “orientado a projeto”, proposto por (Garzotto et al., 1995), considera o seguinte subconjunto de critérios de avaliação: **riqueza, facilidade, auto-evidência, previsibilidade, legibilidade, consistência e reuso**.

### 3. Atributos dos *Links* da *Web* para avaliação

Para cada um dos critérios de (Hatzimanikatis et al., 1995) e (Garzotto et al., 1995), apresentados na seção anterior, por sua vez podem ser atribuídas métricas que auxiliem à sua avaliação. De fato, a qualidade pode ser vista como um parâmetro (ou um indicativo) que pode ser estimado por meio de métricas (de um atributo de interesse) (Jalote, 1997). No entanto, as métricas não são fáceis de se estabelecer, uma vez que são variáveis dependentes de características intrínsecas ao software e seu processo. Dessa forma, para a abordagem de avaliação de *links* da *Web* que investigamos, as métricas estão sendo estudadas e analisadas quais as mais representativas. Neste estudo estamos considerando também quais métricas podem ser implementadas a partir das informações fornecidas pelo sistema LiOS (*Link-Oriented System*) (Fortes, 1996). Este sistema tem como objetivo principal verificar indicativos de qualidade de hiperdocumentos da *Web*, quanto aos critérios de reuso e consistência de *links*. LiOS fornece uma lista de todos os *links* extraídos sendo que um *link* é um objeto composto por nó-fonte, âncora e nó-destino. Recentemente, uma nova versão deste sistema foi construída, contando com o suporte de SGBD para controlar as versões das informações coletadas dos *sites*; o sistema foi denominado DB-LiOS (*DataBase - Link Oriented System*) (Seraphim & Fortes, 2000).

Para que os *links* de hiperdocumentos da *Web* sejam avaliados orientados por meio de métricas significativas, um dos passos mais importantes deste trabalho será a aplicação prática das métricas estudadas para análise de páginas. Entre as métricas que podem vir a ser coletadas a partir das informações fornecidas pelo DB-LiOS podemos ter: compactação, estratificação, número de *links* que direcionam/chegam numa certa localidade (URL/diretório) na *Web* e outras que estão sendo definidas. Ao aplicar essas métricas teremos medidas objetivas. O nosso objetivo é verificar quão boas são essas métricas para realizar a análise estrutural e das páginas de um hiperdocumento da *Web*. Para atingir esse objetivo a nossa proposta inicial é considerar a opinião de um usuário especialista, o qual deverá classificar páginas, de um determinado domínio, em diversas categorias de qualidade. Considerando, para cada página de um site, os valores obtidos através das diversas métricas e a classificação do especialista, tem-se a informação apropriada para utilizar sistemas de aprendizado de máquina proposicionais, a fim de verificar a importância dessas métricas na classificação realizada pelo especialista. Com isso, teremos resultados de um experimento que deve mostrar a relevância de determinadas métricas durante a avaliação da qualidade de hiperdocumentos.

Um outro passo deste trabalho é a classificação de *links*. Uma consideração importante neste passo é o enfoque dado às âncoras, que são os pequenos textos ou imagens apresentados ao usuário. Essas âncoras deverão ser investigadas para determinar quais os possíveis atributos podem ser considerados na classificação. Neste sentido, um estudo da sintaxe do elemento *link* é relevante para verificar todos os atributos que podem ser utilizados. Além disso, é importante definir tipos de âncoras que serão relevantes na classificação e para tanto um grande volume de âncoras de um determinado domínio deve ser analisado para que os tipos mais comuns e representativos sejam identificados. Para esta classificação pretendemos utilizar alguns sistemas de aprendizado de máquina relacionais (Lavrac & Džeroski, 1994; Slattery & Craven, 1998).

#### 4. Considerações Finais

A avaliação da qualidade de hiperdocumentos é uma atividade importante e o auxílio de ferramentas automáticas para análise das características inerentes ao hiperdocumento é um suporte fundamental para tal atividade. Este trabalho deve auxiliar a avaliação de hiperdocumentos da *Web* sugerindo possíveis modificações na sua estrutura, subsidiada pela análise de seus *links* a ser realizada com o suporte de classificação e métricas de *links*.

#### Referências Bibliográficas

- (Blustein et al., 1997) BLUSTEIN, J.; WEBBER, R. E.; TAGUE-SUTCLIFFE, J.: "Methods for evaluating the quality of hypertext links" *Information Processing & Management*, v.33, n.2, p. 225-271, 1997.
- (Botafofo & Shneiderman, 1992) BOTAFOGO, R. A.; SHNEIDERMAN, B.: "Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics" *ACM Transactions on Information Systems*, v.10, n.2, p.142-80, April, 1992.
- (Brown, 1990) BROWN, P. J.: Assessing the quality of Hypertext documents In Rizs, A.; Stritz, N.; André, J.; eds. "Hypertext: Concepts, System and Applications" Cambridge, Great Britain, Cambridge University Press, 1990. Proceedings of ECHT'90, Versailles, France, November 1990./ p1-12.
- (Ebert, 1996) EBERT, C. *Software Quality Journal* 5, (1996), p. 255-272.
- (Fortes, 1996) FORTES, R. P. M.: "Análise e Avaliação de Hiperdocumentos: uma abordagem baseada na Representação Estrutural". Tese de Doutorado, IFSC-USP, São Carlos - SP. 30 de agosto de 1996. 179p.
- (Fortes & Nicoletti, 1997) FORTES, R. P. M.; NICOLETTI, M. C.: "A Family of Link Based Metrics for the Evaluation of Web Documents" *SIGLINK Bulletin*, v.6, n.3, p.21-23, October 1997.
- (Garzotto et al., 1995) GARZOTTO, F.; MAINETTI, L.; PAOLINI, P.: "Hypermedia Design, Analysis, and Evaluation Issues" *Comm. of the ACM*, v.38, n.8, p.74-86, August 1995.
- (Hatzimanikatis et al, 1995) HATZIMANIKATIS, A. E.; TSALIDIS, C. T.; CHRISTODOULAKIS, D.: "Measuring the Readability and Maintainability of Hyperdocuments" *Software Maintenance: Research and Practice*, v.7, p.77-90, 1995.
- (Jalote, 1997) JALOTE, P. *An Integrated Approach to Software Engineering*. 2<sup>nd</sup> Edition, Springer, USA, 1997.
- (Lavrac & Džeroski, 1994) LAVRAC, N.; DŽEROSKI, S. *Inductive Logic Programming: Techniques and Applications*. Ellis Horwood, 1994.
- (Rivlin et al., 1994) RIVLIN E.; BOTAFOGO, R.; SHNEIDERMAN, B.: "Navigating in Hyperspace: Designing a Structure-based toolbox" *Comm. of the ACM*, v.37, n.2, p.87-96, February, 1994.
- (Seraphim & Fortes, 2000) SERAPHIM, E.; FORTES, R. P.M. - "DB-LiOS: Suporte Automático à Avaliação da Consistência de *Links* em *WWW*" - In: XX Congresso Nacional da SBC, XXVII SEMISH, *Anais*. Curitiba-PR, julho de 2000.
- (Slattery & Craven, 1998) SLATTERY, S.; CRAVEN, M.: "Combining statistical and relational methods for learning in hypertext domains" In *Proceedings of the 8th International Conference on Inductive Logic Programming*, Springer-Verlag, 1998.