

Relações Temporais para Indexação e Busca de Texto na WEB: Estudo de Caso para um Ambiente Acadêmico

Niccolas Vidal , Leila Lisiane Rossi, José Valdeni de Lima

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)

Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brasil

nick, llrossi, valdeni@inf.ufrgs.br

***Abstract** This paper offers an overview of an extension of the alocLISA environment (Academic Literature Organized with Computer Assistance) with the goal of retrieving valid temporal texts. This extension based on Allen's temporal model will help authors better organize themselves by finding relevant texts considering temporal aspects.*

***Resumo** Este artigo propõe a extensão do ambiente alocLISA (Ambiente de Localização e de Organização de Conteúdo Acadêmico) visando a busca de texto e sua validade temporal. Esta extensão, baseada no modelo de Allen, permitirá os autores recuperarem textos temporalmente relevantes facilitando a organização de suas tarefas.*

1. Introdução

Dada a característica pervasiva da Web, ela se tornou um ambiente favorável para a divulgação de textos. Porém, são poucos os sistemas de recuperação de informação que levam em consideração o aspecto temporal da validade do texto. O objetivo deste trabalho é propor uma extensão do ambiente alocLISA (Ambiente de Localização e de Organização de Conteúdo Acadêmico), cujo protótipo foi implementado na linguagem PHP, para permitir a busca de textos levando em conta a validade temporal dos mesmos.

A estrutura do trabalho é a seguinte: primeiramente é feita uma introdução ao modelo temporal baseado em intervalos proposto por Allen. Depois, a extensão do ambiente alocLisa é descrita. Finalmente, seguem a conclusão e as perspectivas para trabalhos futuros.

2. Modelo de Allen

O modelo de [Allen 1983] serve de base para a definição das operações necessárias para a busca de textos dentro de certos intervalos temporais referentes à validade do texto. Dois intervalos podem ser relacionados de treze maneiras diferentes, existindo sete relações básicas como ilustrado na figura a seguir. Seis dos sete casos correspondem às relações inversas, como por exemplo *A before B* e *B before A*. No entanto, a relação *equals* não tem inversa, uma vez que *A equals B* é a mesma relação que *B equals A*.

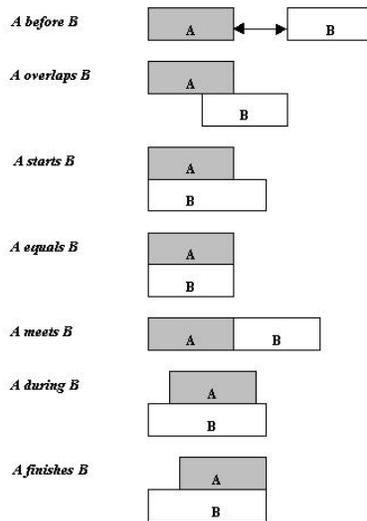


Figura 1 Relações Temporais entre dois objetos

Neste trabalho, foram eleitas duas relações temporais de Allen (*A equals B* e *A during B*) como essenciais para a definição das operações com os intervalos de tempo visando identificar os textos válidos para o planejamento das ações futuras dos autores. A seguir é mostrado um exemplo da aplicação destas operações na submissão de artigos.

Na operação *A equals B*, considera-se *A* como o “intervalo de validade do texto”, e *B* como o “intervalo de validade possível da ação”; ou seja, o autor poderá submeter um artigo em um determinado congresso no “intervalo de validade do texto”. Em outras palavras, significa que se a relação *A equals B* é verdadeira, a ação do aluno é possível e válida. Na operação *A during B*, o autor poderá submeter o artigo durante o intervalo de tempo *A*, podendo ter mais tempo para preparar a ação e executá-la no intervalo de validade do texto. Existem outras maneiras de utilização das relações de Allen no domínio de aplicação alvo deste trabalho que não serão discutidas por falta de espaço, mas que não prejudicam a compreensão do presente artigo.

3. Extensão temporal do alocaLISA

O alocaLISA (*Ambiente de Localização e de Organização de Conteúdo Acadêmico*) é um ambiente controlado de publicação, classificação e avaliação de textos acadêmicos que utiliza três importantes técnicas de ranking: com base na classificação especialista [Bharat 1999], na informação de utilização [Bharat 1999] e na conectividade [Brin 1998]. As principais funcionalidades deste ambiente são a seguir exemplificadas:

- Um professor pode criar um canal para gerenciar uma determinada disciplina ou um determinado grupo de pesquisa;

- Alunos e professores podem se inscrever (ou serem automaticamente inscritos) neste canal;
- Ao acessar o ambiente, somente os canais aos quais o usuário pertence serão mostrados, sendo os outros devidamente filtrados.
- Uma vez inscritos em um canal, alunos e professores podem contribuir enviando textos;
- O professor que gerencia o canal, recebendo os textos, analisa a qualidade e originalidade destes e permite ou não a sua publicação;
- Alunos e professores inscritos neste canal podem enviar comentários e avaliar os textos.
- A cada usuário está associado um fator identificando o poder de participação deste dentro de um determinado canal que depende da qualidade e da quantidade de textos por ele já publicados. Este fator é levado em consideração quando um texto é avaliado.
- As informações dos textos publicados neste ambiente controlado são propagadas para classificar os textos referenciados, mesmo aqueles que não pertencem ao ambiente.

No alocLISA, cada canal trata de uma área específica e é gerenciado por um professor. O benefício deste ambiente é o incentivo à publicação de artigos. Para facilitar ainda mais o processo de publicação, a extensão temporal proposta visa identificar congressos relevantes para os quais os integrantes de cada canal podem submeter seus trabalhos.

Um aspecto importante para determinar se um congresso é relevante ou não é justamente a data limite para a submissão de trabalhos. Outro aspecto seria a data de realização do congresso. Um congresso cuja data limite de submissão já passou ou cuja data de realização coincide com alguma outra atividade importante já agendada não é muito interessante. Ou seja, o fator temporal é extremamente importante quando se quer determinar a relevância de um congresso. Os integrantes poderiam ainda fazer uma busca indicando os prazos que lhes mais favorecem.

Como o assunto tratado dentro de cada canal é bem específico, pode-se utilizar os artigos publicados neste canal para determinar que tipo de congressos são de interesse para os integrantes. Uma consulta aos trabalhos publicados anteriormente poderá servir de auxílio para a criação de um ranking dos congressos. Os congressos onde os melhores artigos foram publicados seriam melhor classificados.

4. Trabalhos Relacionados

O modelo temporal de Allen é comumente utilizado para definir uma apresentação multimídia sincronizada como pode ser observado em [Rousseau 1999]. A extensão do alocLISA, no entanto, preocupa-se em utilizar o modelo de Allen para a validação temporal de textos.

O motor de busca [Google 2003] permite realizar consultas considerando a data de atualização dos textos (e.g. exibir os textos atualizados nos últimos três meses). Esta data está associada à data de gravação do arquivo e não tem nenhuma relação com o texto do próprio arquivo.

Em [Allan 2001] uma versão antiga e uma versão nova de um texto são comparadas para identificar possíveis modificações (*time-based summarization*). Porém, a validade temporal de um texto continua não sendo tratada, sendo que o que está em análise é somente a variação deste texto ao longo do tempo.

5. Conclusão

Os trabalhos pesquisados na área e referenciados no decorrer deste artigo não consideram o aspecto temporal para a validação de textos na Web. A extensão do alocLISA tem como objetivo proporcionar esse diferencial nas pesquisas e como estudo de caso ela foi utilizada em um ambiente acadêmico para facilitar o planejamento dentro de um intervalo de tempo futuro.

Como trabalho futuro é fundamental aplicar técnicas de tratamento da linguagem natural para a identificação do “intervalo de validade temporal do texto” de forma automática. O protótipo atual identifica o referido intervalo de forma manual.

A ferramenta de busca alocLISA está sendo melhorada através de técnicas de consulta agrupadas e a inclusão do fator tempo. A intenção é disponibilizá-la para o uso da comunidade acadêmica da UFRGS no âmbito de uma disciplina, em um primeiro momento, para em seguida disponibilizá-la de forma mais ampla em outras instituições e/ou órgãos de interesse.

6. Referências

- Allan, J., Gupta, R., Khandelwal, V. Temporal summaries of news topics. In Research and Development in Information Retrieval, pages 10-18, 2001.
- Allen, J. Maintaining Knowledge about Temporal Intervals, Communications of the ACM, vol.26, nº 11, 1983.
- Bharat, Krishna; Mihaila, George. Hilltop: A Search Engine based on Expert Documents, 1999. Disponível em <http://www.cs.toronto.edu/~georgem/hilltop/>.
- Brin, Sergey; Page, Lawrence. The Anatomy of a Large-Scale Hypertextual Web Search Engine , 1998. <http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm>
- Google, 2003. Disponível em <http://www.google.com/>
- Rousseau, Franck; GARCIA-MÁCIAS, J. Antonio; LIMA, José Valdeni de Lima; DUDA, Andrzej. User adaptable multimedia presentations for the WWW, 1999. Disponível em <http://drakkar.imag.fr/publications/www8/>.
- Vidal, Níckolas; Feijó, Diego; Lima, José Valdeni. alocLISA: Ambiente de Localização e Organização de Conteúdo Acadêmico. Relatório de Pesquisa UFRGS, 2001.

Agradecimentos

Este trabalho contou com o apoio financeiro do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).