



A recuperação da informação em texto jornalístico: estudo de caso do Núcleo de Estudos da Violência

Iraci Oliveira Rodrigues e Nair Yumiko Kobashi

Resumo: Apresentação dos resultados de pesquisa sobre acesso ao conteúdo do Banco Digital do Núcleo de Estudos da Violência, da Universidade de São Paulo (NEV/USP), que atualmente armazena 40.000 notícias de Imprensa para uso em pesquisas sobre as Graves Violações dos Direitos Humanos no Brasil. O NEV/USP monitora os casos de Violência Policial, Execução Sumária e Linchamento publicadas em veículos da imprensa. Armazena também recortes de jornais sobre os temas: Justiça, Outros Crimes da Polícia, Sistema Prisional, Violência contra a Mulher, Pena de Morte, Menoridade Penal, Segurança Pública, Criminalidade, Milícias, Desarmamento, entre outros. Nesta pesquisa foi inicialmente traçado um breve histórico da cobertura jornalística sobre violência no Brasil, marcada por duas fases, sendo a primeira (1950-1990) caracterizada pelo sensacionalismo das notícias e proximidade da imprensa com a polícia, e, a segunda fase, que teve início nos anos 90 e vigente até os dias de hoje, se caracteriza pela ampliação da cobertura jornalística sobre violência. Inclui novas temáticas como Segurança Pública e traz para discussão outros atores, como autoridades da Segurança Pública, pesquisadores e organizações da sociedade civil. Apesar de caracterizada por duas fases, a cobertura jornalística sobre violência ainda hoje conta com momentos de retorno à primeira fase.

Para identificar os problemas de busca de conteúdos armazenados no Banco Digital, realizamos testes de busca e recuperação que foram analisados quantitativa e qualitativamente. Constatou-se que o uso da linguagem natural como única alternativa nas buscas, em Bancos digitais, dessa envergadura, interfere negativamente na recuperação da informação. A partir dos resultados dos testes foram propostas melhorias na qualidade do Banco Digital, com base em referencial teórico da Ciência da Informação. Propõe-se a utilização de um sistema híbrido de organização de conteúdos que concilie linguagem natural e vocabulário controlado. Essa proposta vem no sentido de promover melhor disponibilização das informações para os pesquisadores da instituição e para os cidadãos. O trabalho discute, por fim, a importância social e política da manutenção dos arquivos de textos jornalísticos por instituições de pesquisa e sua disponibilização para cidadãos e pesquisadores. Mostra também as contribuições dos profissionais bibliotecários neste processo. Destaca, nessa medida, a importância da

adoção de princípios e métodos consistentes de organização da informação, tendo como foco o uso de vocabulários controlados para *taggear* (indexar) de forma consistente a informação jornalística.