

Um sistema de disseminação seletiva de informações para o Congresso Nacional

Perola Cardoso Raulino

Biblioteca

Senado Federal

e

Departamento de Biblioteconomia

Faculdade de Estudos Sociais Aplicados

Universidade de Brasília

Brasília, DF

Resumo — A instalação do Centro de Processamento de Dados do Senado Federal, possibilitando a implantação de um sistema de indexação, armazenagem e recuperação de informações, ensejará uma utilização melhor e mais intensa dos diferentes tipos de informações necessárias ao processo decisório. No âmbito desse sistema, haverá um subsistema de disseminação seletiva de informações, cujo projeto é aqui descrito.

Introdução

“Com o avanço da técnica e com as sociedades massificadas, cada vez se torna mais difícil governar. Uma febre de leis vem tomando conta de nossos destinos. Fazer leis virou uma técnica que se complica. Os políticos de todo o mundo vêem aumentadas as dificuldades em tomar decisões. Eles, além de suas ocupações normais, têm que viver numa constante atualização, ameaçados de naufragar no mar de informações e inovações” (3).

Essa citação descreve bem a situação com que se defronta o Congresso Nacional, assoberbado com uma carga de trabalho sempre crescente, que é uma decorrência do próprio desenvolvimento do país. O problema da falta de informações, selecionadas e correntes, que pudessem ajudar os membros do Congresso em suas tarefas legislativas, já foi por eles percebido muitas vezes. Várias soluções foram tentadas para isso: criação de assessorias legislativas, contratação de especialistas em assuntos parlamentares, ou aperfeiçoamento dos órgãos de informação, sem que se chegasse a um resultado satisfatório.

Versão resumida da dissertação submetida como requisito parcial para a obtenção do grau de Mestre em Biblioteconomia e Documentação pelo Instituto Brasileiro de Bibliografia e Documentação/Universidade Federal do Rio de Janeiro, em abril de 1972.

As necessidades do Congresso em matéria de informação são ilimitadas e muito diversificadas, exigindo a utilização de inúmeras fontes para a realização de pesquisas bibliográficas e legislativas. Além das necessidades individuais de cada senador ou deputado, as comissões do Congresso, que têm por finalidade “estudar os assuntos submetidos a seu exame, realizar estudos e levantamentos sobre problemas de interesse nacional compreendidos no âmbito de suas atribuições, acompanhando a execução dos planos e programas administrativos adotados pelo Poder Executivo em todo o território nacional” (1), também exigem informações corretas e atualizadas para bem desempenhar suas funções.

Por outro lado, os métodos convencionais de pesquisa bibliográfica já não atendem ao crescimento do volume de informações e requerem um tempo precioso, sem que se tenha a certeza de que todos os documentos pertinentes serão encontrados. Por isso, é preciso empregar processos não-convencionais para se ter acesso a essa massa de informações.

Estabelecimento dos perfis de usuários

Um sistema de disseminação seletiva de informações (DSI) é um dos meios de enfrentar a situação acima descrita e incrementar o uso da informação. De acordo com o sistema proposto por Luhn (2), a DSI é, basicamente, um método para fornecer, em grande escala, um serviço de notificação corrente para indivíduos ou grupos. O usuário fornece os detalhes dos assuntos que lhe interessam e os conceitos respectivos são traduzidos para os termos de indexação usados no sistema, formando assim o perfil do usuário. Este perfil será incluído no arquivo que, no computador, lhe corresponde. Outro arquivo será reservado na máquina para os dados referentes aos documentos indexados. A comparação entre esses dois arquivos mostrará quando ocorre uma coincidência entre os descritores representativos do perfil do usuário e os descritores que representam o conteúdo dos documentos. Havendo a coincidência, o computador emitirá uma notificação, contendo os elementos que permitirão identificar o usuário e o documento indexado.

A parte mais trabalhosa, na implantação de um sistema de DSI, talvez seja o estabelecimento dos perfis de interesses dos usuários, que podem ser tanto individuais quanto de grupo. Várias são as técnicas empregadas para se fazer esse levantamento, sendo que a mais comum, além da entrevista, é fornecer aos prováveis usuários resumos ou textos de documentos, solicitando-lhes que assinalem os assuntos que lhes interessam. Outra técnica utilizada é a observação dos seus hábitos de leitura e pesquisa, a qual foi por nós adotada, tendo em vista a diversidade e complexidade dos temas debatidos nas duas casas do Congresso Nacional.

Elaboramos um questionário, que foi distribuído a 66 senadores, e que se baseou nos assuntos que são pesquisados com maior frequência nas duas

bibliotecas do Congresso. Para o levantamento desses assuntos foram consultadas várias fontes: arquivos de pesquisas solicitadas pelos congressistas; arquivos de bibliografias compiladas, a pedido, ou por iniciativa da biblioteca; estatísticas diárias; e os projetos de leis em tramitação. Também foram ouvidos os bibliotecários de Referência do Senado e da Câmara.

Explicamos ainda aos senadores, em entrevista pessoal, que poderiam acrescentar outros temas que não constassem da relação de assuntos incluída no questionário e sobre os quais gostariam de receber informações relevantes e selecionadas. Em alguns campos, muito simples, foi pedida a especificação da área de interesse.

O perfil de interesse de cada usuário será vinculado aos seus dados de identificação pessoal, inclusive o seu nome estará entre os descritores, e armazenado em memória de máquina. Durante a implantação do sistema, os perfis sofrerão reajustes e modificações, a fim de se evitar a recuperação de material irrelevante, proveniente de assunto idêntico, mas tratado em nível diferente.

O total de questionários devolvidos foi de 49, sendo que 50 novos temas foram sugeridos, além do total de 154 incluídos de antemão no questionário.

Além do perfil individual, o perfil de grupo, abrangendo todos os membros das várias comissões permanentes, fornecerá aos componentes dessas comissões as informações que lhes servirão de auxílio na elaboração de pareceres e emendas.

As comissões permanentes do Senado são: Agricultura, Assuntos Regionais, Constituição e Justiça, Distrito Federal, Economia, Educação e Cultura, Finanças, Legislação Social, Minas e Energia, Relações Exteriores, Saúde, Segurança Nacional, Serviço Público Civil, Transportes, Comunicações e Obras Públicas. Cada uma possui atribuições específicas e o número de seus membros varia de cinco a quinze.

No perfil de grupo, as atribuições de cada comissão, com todas suas relações, serão os assuntos que receberão descritores e irão formar o perfil de grupo.

Pesquisa dos hábitos de leitura

No questionário foi incluída, também, uma relação dos periódicos mais utilizados pela biblioteca para a localização de informações solicitadas pelos congressistas. Essa relação resultou de um levantamento feito nas duas bibliotecas do Congresso, tomando-se por base os periódicos mais utilizados pelos bibliotecários para atender à demanda de informações.

Solicitou-se aos senadores que assinalassem os títulos de periódicos que costumavam compulsar, para localização de informações úteis, podendo ainda acrescentar outros títulos que não tivessem sido relacionados.

Nos 49 questionários devolvidos, observamos que somente 13 títulos novos foram incluídos sendo que 9 usuários não assinalaram nenhum título como sendo de consulta regular (Quadros 1 e 2). Este último fato comprova a necessidade de um serviço de informação mais agressivo e atualizado.

Constatamos que o conceito de sistema de DSI era desconhecido por quase todos. Após uma explicação sobre seu funcionamento e os benefícios que traria, alguns usuários o aceitaram com grande entusiasmo.

Entrada das informações

As informações a serem processadas pelo computador do Centro de Processamento de Dados do Senado Federal (PRODASEN), incluindo os descritores representativos dos assuntos dos documentos, serão registradas em quatro diferentes arquivos ou bases de dados (*data bases*): a) arquivo de livros; b) arquivo de leis; c) arquivo de periódicos; d) arquivo de controle de projetos e comissões.

O controle desses arquivos será entregue ao STAIRS (*Storage and Information Retrieval System*), que permitirá criar, estender e manter bases de dados por meio de programas utilitários que funcionam mediante processamento em lotes.

Arquivos de livros

O processamento de livros na Biblioteca do Senado segue as rotinas usuais, adotando-se cabeçalhos de assuntos para representação de seu conteúdo.

Os dados catalográficos, mais um número de identificação, darão entrada no computador através de cartões perfurados, em lotes, sendo depois transferidos para fitas magnéticas e, finalmente, após sua correção, para discos magnéticos.

O número de identificação, número de chamada, local, editor, data, paginação e ilustração ocupam campos fixos, e as outras informações ocupam campos variáveis. Quando for processada a comparação entre os descritores dos livros e os dos perfis, o programa permitirá a localização direta do campo, evitando a leitura completa do texto.

Será emitida uma notificação com os elementos identificadores do documento e também os descritores, que indicarão ao usuário os assuntos dos documentos referenciados.

Os programas a serem utilizados para o subsistema de referência bibliográfica serão os mesmos aplicados à DSI.

Arquivo de leis

As leis, decretos e decretos-leis terão seu registro no computador pelo número, data da publicação, ementa e possíveis alterações e republicações. Toda

QUADRO 1. *Periódicos brasileiros mais consultados pelos senadores, de acordo com as respostas dadas ao questionário*

Título do periódico	Número de consulentes
Conjuntura Econômica	30
Revista Brasileira de Estudos Políticos	29
Problemas Brasileiros	26
Revista de Informação Legislativa	23
Revista de Ciência Política	21
Revista Brasileira de Economia	18
Educação Hoje	16
Digesto Econômico	15
Revista Brasileira de Política Internacional	14
Indústria e Produtividade	14
Revista de Administração Pública	12
Mundo Econômico	12
LTR — Legislação do Trabalho	12
Revista Forense	12
Revista Econômica	11
Revista de Direito Público	11
Revista de Administração de Empresas	10
Revista do Serviço Público	10
Revista de Direito Administrativo	10
Revista de Administração Municipal	9
Revista de Finanças Públicas	9
Jurídico	8
Bolsa	8
Revista dos Transportes	8
Boletim Eleitoral	7
Carta Mensal	6
Revista de Política e de Administração Fiscal	5
Revista da Ordem dos Advogados do Brasil	5
Revista de Direito Mercantil e Financeiro	5
Revista Bancária Brasileira	4
Revista Paranaense de Desenvolvimento	2

QUADRO 2. Periódicos estrangeiros mais consultados pelos senadores, de acordo com as respostas dadas ao questionário

Título do periódico	Número de consultentes
Revista de Estudios Políticos (Madrid)	13
Actualités Economiques (Paris)	10
Economist (London)	10
Revue Politique et Parlementaire (Paris)	8
Il Politico (Pavia)	6
American Economic Review (Providence)	6
American Political Science Review (Washington)	5
Seguridad Social (México)	5
Revue du Droit Publique et de la Science Politique (Paris)	4
Revista Internacional del Trabajo (Genève)	4
Parliamentary Debates (London)	4
Parliamentary Affairs (London)	4
Journal of Political Economy (Chicago)	4
Económica (London)	4
Foreign Affairs (New York)	3
Revista de Derecho Financiero y de Hacienda Pública (Madrid)	3
Economic Journal (Cambridge)	3
Constitutional and Parliamentary Information (Genève)	3
Comercio Exterior (México)	3
Revue Internationale de Droit des Gens (Toulon)	2
American Journal of International Law (Washington)	2
Droit Social (Paris)	1
Economia Internazionale (Genoa)	1
Notes et Etudes Documentaires (Paris)	1

essa legislação será referenciada também por assunto, empregando o mesmo vocabulário de indexação utilizado para o arquivo de livros.

No processo de recuperação de textos legislativos, a pesquisa dos descritores se fará não só nos cabeçalhos de assuntos, mas também nas ementas.

A notificação a ser emitida conterá todas as informações que tiverem sido armazenadas no computador a respeito de cada documento legal.

Arquivo de periódicos

A todos os periódicos recebidos pela Biblioteca serão atribuídos códigos de seis algarismos que permitirão intercalações, pois a sua seqüência numérica estará de acordo com a seqüência alfabética dos títulos dos periódicos.

Apenas o código será utilizado para a entrada em máquina das referências dos artigos, mas ele será substituído pelo título completo do periódico quando o computador imprimir a notificação.

As referências dos artigos selecionados dos periódicos constarão de código do periódico, autor, título do artigo, título do periódico, números do volume e do fascículo, e a data. Naturalmente, serão atribuídos aos artigos os descritores representativos de seu conteúdo. Na fase inicial de implantação do sistema, serão utilizados apenas os descritores, mas, numa segunda etapa, serão preparados resumos dos artigos, os quais também farão parte das notificações emitidas pelo computador.

Será feita também a DSI de periódicos, nos moldes já descritos para a legislação, e a listagem de todos os artigos referenciados, dispostos sob grandes cabeçalhos de assuntos, servirá para a preparação de um boletim da biblioteca, a ser distribuído aos interessados.

Arquivo de controle de projetos e comissões

Este arquivo controla o andamento de todas as proposições apresentadas ao Congresso e permite ao usuário tomar conhecimento de sua tramitação e de suas características. Prevê também a possibilidade de o legislador identificar quais os projetos rejeitados e arquivados.

Serão estabelecidos, pelo menos, dois documentos para cada matéria: 1) Dados básicos; 2) Ações legislativas. Dados básicos: identificação, números, ementa, assuntos, autoria, legislação, publicações. Ações legislativas: identificação, ação administrativa, ação legislativa, *header*. Todas as informações serão perfuradas em cartões e darão entrada em lotes no computador, sendo transferidas para discos magnéticos. Futuramente, a alimentação será em linha.

Como o nome do próprio usuário se acha entre os descritores, toda vez que o sistema localizar uma comparação nos textos completos desse arquivo

emitirá notificações informando ao interessado sobre a tramitação dos projetos, emendas, pareceres, etc. de que seja o autor.

Disseminação

Os perfis de interesse dos usuários, previamente registrados na biblioteca de consultas, serão utilizados para a disseminação seletiva dos novos itens incluídos no subsistema. Em função dos perfis e da data da última distribuição serão impressos os itens de interesse de cada usuário.

Os cartões de notificação gerados pelo sistema, na fase inicial, serão conferidos manualmente, para verificação da referência selecionada, se está dentro do campo do perfil registrado, revelando, em certos casos, a necessidade de modificação do perfil. Esses cartões poderão ser coloridos, para chamar a atenção, e terão uma parte destacável que será devolvida pelo usuário ao órgão competente. Os quesitos que poderão constar da notificação dos artigos de periódicos e que serão aproveitados para reajuste do perfil são: Artigo de interesse, envie texto completo; Artigo de possível interesse futuro; Artigo de interesse, mas já foi visto; Artigo sem interesse; Artigo em língua desconhecida, gostaria de receber a tradução.

As respostas aos itens acima indicam se as referências selecionadas de acordo com o perfil são relevantes ou não para os interesses do usuário, proporcionando um meio de avaliar a eficácia do sistema e melhorar o índice percentual de relevância.

Processo de comparação

Há vários critérios que podem ser aplicados, num sistema de DSI, para selecionar somente os documentos relevantes destinados a cada usuário. Esta precisão faz parte de qualquer sistema bem estruturado de recuperação de informações. O sistema proposto será uma combinação de pesquisa livre em todo o texto ou uma busca nos campos de descritores que obedecem a uma terminologia previamente fixada.

Para melhorar o índice de relevância aplicar-se-á a lógica booleana na recuperação das informações. Serão empregados os seguintes operadores binários:

- 1) a *ADJ* b: a e b são adjacentes e estão nesta ordem na mesma sentença. Exemplo: ADMINISTRAÇÃO ADJ ESTADUAL.
- 2) a *WITH* b: a e b ocorrem na mesma sentença. Exemplo: FILOSOFIA WITH DIREITO; ABUSO WITH PODER ADJ ECONÔMICO.
- 3) a *NOT* b: a ocorre em documentos e b não ocorre nesses documentos. Exemplo: ENERGIA NOT NUCLEAR.

4) a *AND* b: ambos ocorrem no mesmo documento. Exemplo: CUSTO WITH VIDA AND SALÁRIO.

5) a *OR* b: a ou b ocorrem no mesmo documento. Exemplo: PODER ADJ EXECUTIVO OR PODER ADJ LEGISLATIVO.

6) a *SYN* b: palavras ou frases a e b são consideradas como sinônimos durante a pesquisa. Exemplo: PRODUÇÃO AND AGRICULTURA SYN PRODUÇÃO ADJ AGRÍCOLA.

As operações acima têm duas cadeias de entrada, representando os operando a e b e uma cadeia de saídas representando o resultado.

Há, ainda, uma outra modalidade que o sistema permite: as palavras de consulta podem ser mascaradas para pesquisar todas as palavras que começam com um grupo de caracteres, colocando o símbolo \$ ou este símbolo e mais dois dígitos. Exemplo: ECON\$, que permite a recuperação de Economia, Econômico, Economista, Econometria.

Acreditamos, assim, obter um alto índice de documentos relevantes, evitando fornecer ao usuário materiais que não sejam importantes para seus estudos e pesquisas.

Equipamento

Como já foi antes mencionado, será utilizado um computador IBM/370, modelo 155, com memória de 512 K. Sua capacidade, em discos magnéticos, é de 600 milhões de *bytes*. Dispõe de 8 terminais 2740, com teclado e impressora, e 10 terminais 2260, com teclado e vídeo. A entrada de informações poderá ser feita por cartões perfurados e fita magnética. Sua capacidade de impressão é de 1 100 linhas por minuto (132 caracteres por linha). Pode perfurar 300 cartões por minuto.

A linguagem usada será COBOL. As funções do STAIRS, para criação e atualização das bases de dados, são realizados por cinco programas utilitários: *Text File Build Program*, *File Analysis Program*, *Search File Build Program*, *Data Base Merge Program*, e *Synonym Chain Program*.

Organização dos arquivos

Como já vimos, as informações que serão utilizadas no sistema de DSI serão registradas em quatro arquivos diferentes. Haverá, portanto, quatro bases de dados: para os livros, para as leis, para os periódicos, e para o controle de projetos e comissões. Cada uma das bases de dados compõe-se dos seguintes arquivos: *Text Index File*, *Text File*, *Inverted File*, e *Dictionary File*.

Através do programa, que será diferente para cada base de dados, os dados serão transformados para o formato exigido pelo programa principal.

O *Text File* é um arquivo que contém o texto normal dos documentos, numa forma semelhante à que é impressa ou apresentada no terminal de vídeo. O *Text File Index* contém, em campos formatados, o texto extraído do documento e aponta para o início de cada documento no *Text File*. O *Inverted File* contém a cadeia de todas as ocorrências de palavras na base de dados. O *Dictionary File* contém cada palavra diferente existente na base de dados. Junto à palavra há um *pointer* para a cadeia de todas as ocorrências daquela palavra na base de dados. O dicionário também contém *pointers* para os sinônimos.

Os programas de criação e manutenção das bases de dados serão em lote e o computador fará a interligação entre os arquivos mediante o programa. Toda base de dados arquivará seus dados em duas áreas: a dos dados que já foram processados pelo sistema e a dos novos dados que serão consultados quando o computador fizer a próxima comparação dos perfis e informações. Essa técnica eliminará a recuperação dos dados que foram obtidos anteriormente.

Quinzenalmente, as novas informações que deram entrada nas bases de dados serão comparadas com os perfis através de um comando pelo terminal, ou em lotes.

Conclusões

O fator mais importante na avaliação de um sistema de DSI é seu efeito sobre as atividades que o usuário desenvolve. O sistema estará, realmente, sendo de grande valor para os usuários na medida em que aumentar sua criatividade, produtividade e eficiência, evitando duplicação de esforços ou impedindo que informações úteis sejam ignoradas.

Existem diferentes opiniões sobre a melhor maneira de apresentar as informações recuperadas: anexando-lhes descritores, resumos, as primeiras páginas dos artigos ou, simplesmente, os títulos dos documentos. Optamos pelo uso de descritores, mas nada impede que, após a implantação do sistema e havendo o suporte de uma infra-estrutura adequada, possam os documentos ser resumidos, oferecendo-se esse novo serviço aos usuários.

Um sistema ideal de DSI será aquele que, além da notificação corrente individual, fornecer também estatísticas que permitirão tomar certas decisões de grande valia, como: política de aquisição, política de descarte, prioridades de tradução. O sistema deverá contar com os recursos para fornecer ao usuário, imediatamente após a sua requisição, cópias dos textos completos referenciados na notificação.

A revocação (*recall*) e a precisão (*precision*) são as duas maneiras usuais de se medir a eficiência de um sistema de DSI. A revocação é a relação entre os documentos relevantes recuperados e o número total de documen-

tos relevantes existentes na coleção. A precisão é a relação entre os documentos relevantes recuperados e o número total de documentos recuperados.

A precisão pode ser medida com certa facilidade, pois depende das respostas dos usuários, obtidas em diferentes etapas. A revocação, no entanto, é mais difícil de ser calculada, sendo possível apenas chegar-se a um cálculo estimativo.

Acreditamos que um sistema de DSI proporcionará a todos os congressistas possibilidades melhores para o pleno desempenho de suas funções de legislar e fiscalizar a administração, instruir e orientar o público sobre assuntos políticos de interesse nacional.

Abstract

Selective dissemination of information service at the National Congress

Installation of a data processing center at the Federal Senate and the consequent establishment of an information storage and retrieval system will contribute to increasing and improving the use of the different kinds of information which are necessary for decision-making procedures. The project for a subsystem of selective dissemination of information is described.

REFERÊNCIAS

1. BRASIL. Congresso. Senado Federal. *Regimento interno*. Brasília, 1971, p. 39.
2. LUHN, H.P. Selective dissemination of new scientific information with the aid of electronic-processing equipment. *American Documentation* 12 (2) :131-138, Apr. 1961.
3. MARTINS, Rui Nogueira. Técnicas de informação nas instituições democráticas. *Problemas Brasileiros* 9 (97) :7-18, set. 1971.