**Universidade Federal de Rondônia**

**Departamento Acadêmico de Ciência da Computação**

**Estrutura de Dados II**

**Lista 2 – Arquivos**

**Baseada na lista de exercícios da Prof.ª Carolina Prof.ª Maria Cristina, e livro do Folk.**

1. Explique a diferença entre arquivo lógico e arquivo físico.

**Enquanto o Arquivo físico é uma sequencia de bytes armazenada no disco, o Arquivo Lógico é aquele em que há como acessar por meio de programas ou aplicativos.**

1. Descreva as operações fundamentais que podem ser realizadas em um arquivo. Descreva as funções que executam estas operações na linguagem de programação que você usa, e como elas são utilizadas. Por que existem vários modos de abrir um arquivo?

**Abertura, leitura, escrita e fechamento.**

**Fopen, “r”, “w”, fclose. Ex: FILE \*nomedoarquivo =fopen(“texto.txt”, “r”);**

**Porque para cada função que for usar no arquivo existe uma forma de abertura deste.**

1. Quais as funções de um gerenciador de arquivos?

**O gerenciamento de arquivos tem como função: gerenciar espaços alocados, criar e deletar arquivos .exe, criar e deletar diretórios de memória swap, mapear arquivos em disco em massa, fazer a manipulação de arquivos e diretórios.**

1. Explique porque os arquivos abertos devem ser fechados.

**Para que toda alteração e não alteração feita nele seja salva.**

1. Descreva o que acontece quando um arquivo já existente é aberto por um aplicativo (p. ex., um programa) para escrita. E se o mesmo ocorrer com um arquivo não existente?

**Os dados serão adicionados no fim do arquivo se ele já existir, ou um novo arquivo será criado, no caso de arquivo não existente anteriormente**

1. Os sistemas de arquivos permitem definir atributos para controlar o acesso a um arquivo. O que acontece quando um programa tenta abrir um arquivo que tem proteção para leitura? E quando o arquivo tem proteção para escrita?

**Ele abre o programa num modo de segurança somente para leitura.**

1. Muitos sistemas diferenciam os arquivos binários dos arquivos de texto. Qual a diferença entre eles?

**Nos arquivos texto os dados são representados linha a linha. Na prática é uma sequência de bytes representando caracteres. ... Já os arquivos binários são representados por uma sequência de bytes sem o conceito de quebra de linha. Ele armazena o dado literal, ou seja, não são caracteres.**

1. No que consiste a operação de posicionamento (*seeking*) em um arquivo? Qual a sua utilidade? Exemplifique uma situação em que esta operação precisa ser utilizada.

**O seeking consiste em acessar uma parte especifica do arquivo, ele faz a busca e acessa a posição em que se encontra esse arquivo. Se você precisar achar tal string em determinado arquivo, o ponteiro faz essa busca e acessa o local onde essa string está alocada.**

1. Faça um programa que leia os últimos 10 caracteres de um arquivo qualquer e imprima-os na tela.
2. Faça um programa que leia o conteúdo de um arquivo e o escreve na tela.
3. Faça um programa para copiar o conteúdo de um arquivo para outro.
4. Faça um programa que lê um vetor numérico pelo teclado e o escreve em um arquivo.
5. Discuta as diferenças e semelhanças entre a memória principal (RAM) e a memória secundária (os arquivos).

**As memórias primárias (Reais) são aquelas em que o processador endereça diretamente como: RAM, ROM, Registradores e o CACHE. Já as memórias secundárias são aquelas que não podem ser endereçadas diretamente, todas as informações deverão ser mandadas para uma memória intermediária antes (primária).**

1. Como são organizados fisicamente os discos? De que forma os discos armazenam os arquivos? Por que o tamanho real de um arquivo em disco é sempre maior do que o seu tamanho nominal?

**Cada trilha é dividida em trechos regulares denominados setores. Cada setor possui uma capacidade específica de armazenamento (geralmente, de 512 bytes). A gravação e a leitura dos dados são feitas através de minúsculos eletroímãs presentes nas cabeças de leitura do disco rígido. Compostos de uma bobina de fios que envolvem um núcleo de ferro, esses ímãs são extremamente precisos, o que permite a gravação de trilhas medindo centésimos de milímetros. Porque quanto menor for o tamanho de cada cluster, menos espaço é desperdiçado num sistema de armazenamento.**

1. Quais parâmetros são considerados para calcular o tempo de leitura de um arquivo mantido em disco?

**Ao escolher um disco, três parâmetros são absolutamente essenciais: tempo de seek, tempo de latência e taxa de transferência.**

1. Explique o que é um cilindro, e a razão para a organização de arquivos em cilindros.

**Um cilindro nada mais é do que o conjunto de trilhas com o mesmo número nos vários discos. O braço de leitura é uma peça triangular, que pode se mover horizontalmente. Para acessar um dado contido na trilha 982 da face de disco 3, por exemplo, a controladora do disco ativa a cabeça de leitura responsável pelo disco 3 e a seguir, ordena ao braço de leitura que se dirija à trilha correspondente. Não é possível que uma cabeça de leitura esteja na trilha 982, ao mesmo tempo que outra esteja na trilha 5631, por exemplo, justamente por seus movimentos não serem independentes.**

1. Explique o que é um *cluster* e o que é um *extent*.

**Cluster é o nome dado a um sistema que relaciona dois ou mais computadores para que estes trabalhem de maneira conjunta no intuito de processar uma tarefa. Extent é um sistema de arquivos baseado em extensão mais antigo.**

1. O que é a fragmentação de um arquivo no disco? Quais os tipos de fragmentação do arquivo, por que elas ocorrem e quais seus efeitos?

**A fragmentação ocorre quando o sistema operacional não consegue armazenar determinado arquivo de uma vez só. Gravando-o em diferentes regiões. Esse evento é muito comum no sistema operacional Windows, mas acontece também em outras plataformas, só que em uma escala menor. Isso causa uma lentidão maior na leitura e gravação dos mesmos, do que se estivessem armazenados em um único "pedaço".**

1. Discuta as vantagens e as desvantagens de organizar arquivos em blocos de tamanho definido pelo usuário, ao invés de em setores de tamanho fixo.

**A alocação contígua consiste em armazenar um arquivo em blocos sequencialmente dispostos, permitindo ao sistema localizar um arquivo através do endereço do primeiro bloco e da sua extensão em blocos. O acesso é feito de maneira simples, tanto para a forma sequencial quanto para a direta. A utilização de arquivos organizados sequencialmente oferece problemas. Se o novo registro for maior ou menor que o registro que existia anteriormente, os registros adjacentes podem ser destruídos ou se tornarem inacessíveis quando o novo registro for gravado.**

1. Por que os discos são considerados o gargalo de um sistema computacional? Explique como este problema pode ser minimizado.

**Cada vez que um arquivo é salvo, seu sistema operacional o guarda em uma pasta e o disco rígido (o famoso “HD”) memoriza o “caminho” que leva até esse arquivo. Com o tempo, seus downloads, arquivos salvos e dados de navegação armazenados no disco fazem com que ele memorize “caminhos” demais, efetivamente levando mais e mais tempo para executar seus processos. Apagando programas e arquivos que não são necessários, investindo em HDs externos ou pendrives para desafogar mais espaço do disco rígido.**

1. Quais as vantagens e desvantagens de fitas sobre discos com dispositivos de memória secundária?

**Dispositivos de armazenamento secundários oferecem vários benefícios distintos para o seu computador usar, como possuindo a capacidade de armazenar grandes quantidades de informação, como centenas, até mesmo o equivalente a milhares, de livros. Armazenamento secundário também remove os outrora enormes custos para as empresas que foram incorridos para o armazenamento de documentos importantes em armários ou instalações de armazenamento. Além disso, dispositivos de armazenamento secundário são seguros, confiáveis ​​e permanentes. Enquanto os dispositivos de armazenamento eletrônico de RAM são dispositivos de armazenamento rápido, secundárias são mais lentos porque são eletromecânicos. As informações sobre o dispositivo secundário têm de ser localizadas em primeiro lugar, em seguida, copiados e movidos para a memória RAM ou primário, que é considerado como tempo de busca de dados. Além disso, o armazenamento secundário simplesmente fornece armazenamento para o computador enquanto memória primária , apoia a atividade da CPU em curso , armazenando instruções e dados de programas atualmente em execução.**

1. O que são buffers de E/S (ou I/O)? Quais os passos executados para ler um byte do disco de forma que ele possa ser utilizado por um programa?

**Buffer de dados é uma região de memória física utilizada para armazenar temporariamente os dados enquanto eles estão sendo movidos de um lugar para outro. Quando o S.O inicia o atendimento, ele pode ler o bloco do disco a partir do buffer do controlador. Um bloco de byte ou uma palavra é lida no registrador do controlador e armazenada na memória principal.**

1. As aplicações usualmente armazenam as informações em arquivos organizando-as em campos e registros. Explique as diferentes maneiras pelas quais um campo pode ser armazenado em um arquivo para posterior recuperação.

**Os arquivos são armazenados pelo sistema operacional em diferentes dispositivos físicos, como fitas magnéticas, discos magnéticos e discos ópticos. O tipo de dispositivo no qual o arquivo é armazenado deve ser isolado pelo sistema operacional, de forma que exista uma independência entre os arquivos a serem manipulados e o meio de armazenamento.**

1. Explique as diferentes estratégias que podem ser utilizadas para separar um registro de outro. Discuta as vantagens e desvantagens de cada uma delas.

**Os métodos de ordenação se classificam em:**

**Comprimento fixo: Cada campo ocupa no arquivo um tamanho fixo, pré-estabelecido. O espaço alocado (e não usado) aumenta desnecessariamente o tamanho do arquivo.**

**Indicador de comprimento: O tamanho de cada campo é armazenado imediatamente antes da informação. Requer um único byte se o tamanho do campo é inferior a 256 bytes.**

**Delimitadores: delimitador não pode ser um caractere válido. pode ser um caractere ASCII não imprimível.**

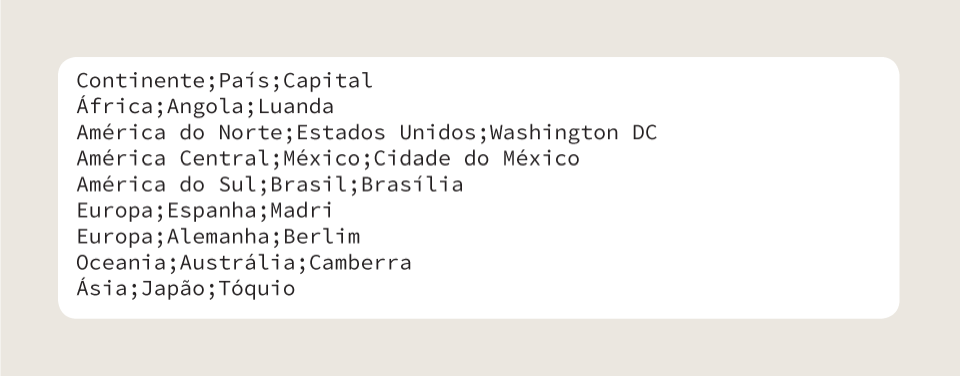
**Uso de tags: possui semântica local, campo fornece informação sobre si próprio 16, permite identificar localmente o conteúdo do arquivo, um campo “perdido” não compromete o arquivo, permite campos existirem ou não. As keywords podem ocupar uma porção significativa do arquivo**

1. Explique o que é fragmentação de campos e registros. Quando e por que ela ocorre?

**A fragmentação dos espaços livres é causada pela criação e eliminação constante de arquivos que com o tempo surgem espaços vagos sem o tamanho suficiente para se alocar novos arquivos.**

1. Se a separação entre registros e campos é feita por delimitadores, quais as restrições para a escolha desses delimitadores? Descreva uma situação que exemplifique sua resposta.

**Um arquivo CSV abriga um sem número de "registros", separados por quebras de linha (cada "registro" permanece numa linha do arquivo) e cada registro possui um ou mais "campos", separados por um delimitador, os mais comuns sendo a vírgula (","), o ponto e vírgula (";") e o caractere "invisível" que surge ao se pressionar a tecla "tab". Arquivos separados por vírgula e ponto e vírgula normalmente recebem a extensão "CSV" e arquivos separados por "tab" a extensão "TSV". Há também bases de dados nesses formatos que recebem a extensão "TXT". Arquivos CSV são simples e funcionam na maior parte das aplicações que lidam com dados estruturados.**



1. O que é gravado no arquivo quando uma struct do C é escrita em um arquivo? Como são armazenados campos que não são strings?

**Gera um arquivo binário. São armazenadas em ponteiro com valor inteiro.**

1. Como um registro é identificado para acesso aleatório? Qual operação permite localizar um registro no arquivo em C?

**Um programa pode acessar cada registro com a mesma rapidez e facilidade que qualquer outro. Nesse sentido, a escolha do dado é arbitrária no sentido de que, independentemente de qual item é procurado, tudo o que é necessário para encontrá-lo é o seu endereço, ou seja, as coordenadas em que ele está localizado, como sua linha e coluna. Por meio de ponteiros.**

1. Explique como é possível melhorar o desempenho de um acesso sequencial a todo o conteúdo de um arquivo. Tal solução também garante um melhor desempenho de uma sequência arbitrária de acessos aleatórios? Discuta.

**Ao atender uma requisição de leitura, são trazidos para o cache mais dados que os solicitados pela requisição; além disso, leituras de dados ainda não solicitados podem ser agendadas em momentos de ociosidade dos discos. Dessa forma, futuras requisições podem ser beneficiadas pela leitura antecipada dos dados. Essa política pode melhorar muito o desempenho de acesso sequencial a arquivos e em outras situações com boa localidade de referências.**

1. Quantas leituras são necessárias, em média, para encontrar um registro em um arquivo com N registros usando a busca sequencial? Quantas leituras são necessárias para identificar que um registro não está no arquivo?

**Serão necessárias N leituras em média. E para um registro que não está no arquivo são necessárias N leituras.**

1. Quais as vantagens e as desvantagens de utilizar arquivos organizados em registros de tamanho fixo?

**Facilidade na pesquisa, indicado para situações nas quais o comprimento dos campos é fixo ou apresenta pouca variação. Espaço alocado (e não usado) aumenta desnecessariamente o tamanho do arquivo (desperdício de espaço de armazenamento), solução inapropriada quando se tem uma grande quantidade de dados com tamanho variável.**

1. É vantajoso manter um arquivo separado para armazenar apenas as chaves e os byte *offsets*, ou RRNs, dos registros no arquivo de dados? Como isto afeta a inserção de um novo registro?

**Sim, pois fica mais fácil a localização do registro, pois ele oferece sua posição relativa. Tamanho dos registros, dependente do tamanho dos campos.**

1. Como um registro pode ser eliminado de um arquivo?

**Por meio da abordagem estática. remoção do registro gera espaço sem dados úteis, “programa compactador” reagrupa os registros e diminui o tamanho do arquivo.**

1. O que são modelos abstratos de dados, para que são utilizados e quais as suas vantagens?

**É mais fácil pensar em dados deste tipo como objetos que representam som, imagens, etc. e que têm a sua própria maneira de serem manipulados. modelo abstrato de dados captura a noção de que o dado não precisa ser visto da forma como está armazenado, ou seja, permite uma visão dos dados orientada à aplicação, e não ao meio no qual eles estão armazenados.**

1. Por que é interessante utilizar cabeçalhos nos arquivos?

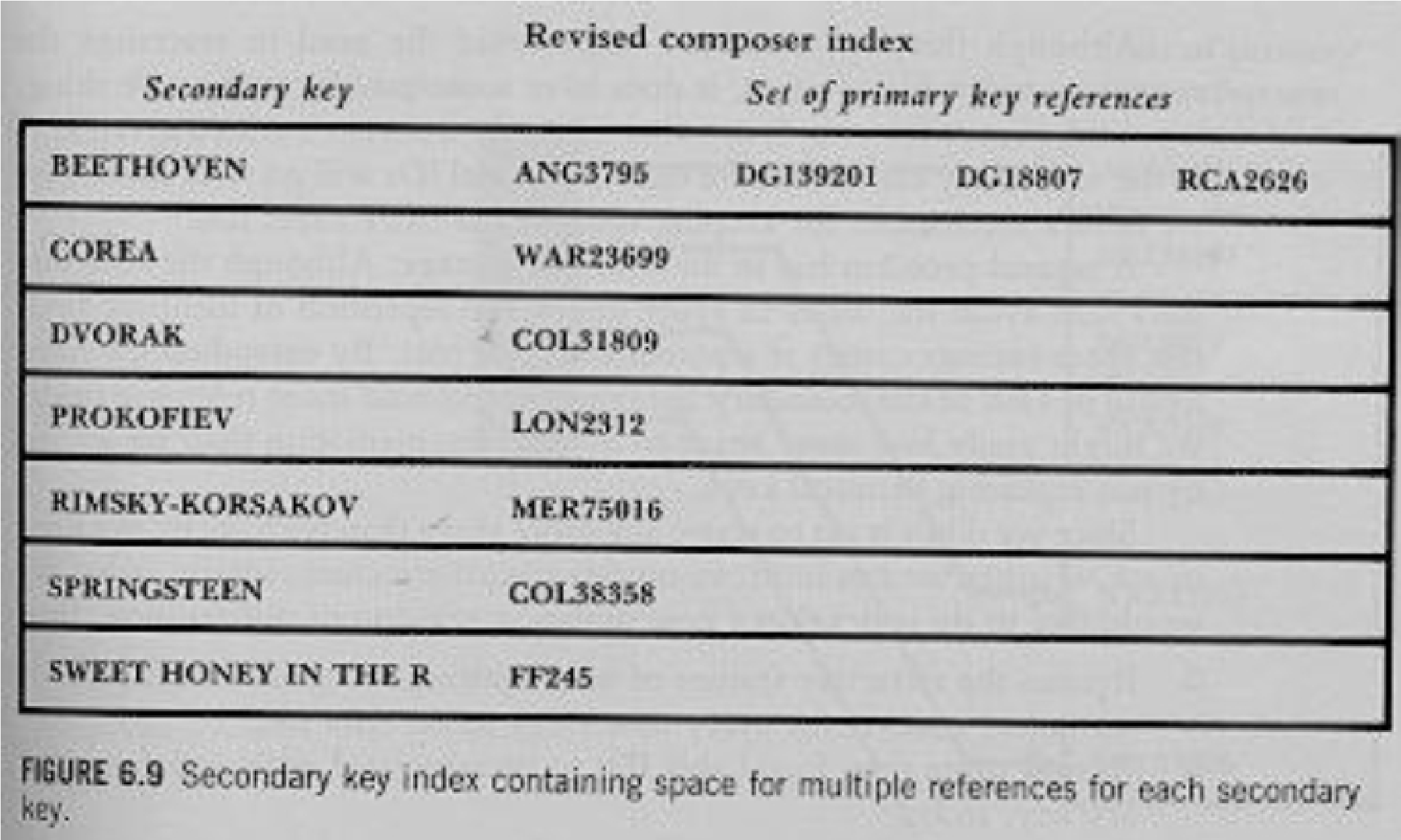
**Porque um arquivo cabeçalho ou arquivos de cabeçalho é um arquivo que permite que os programadores separem certos elementos de um código fonte de um programa em arquivos reutilizáveis. Arquivos de cabeçalho normalmente contêm declarações de envio de classes, sub-rotinas, variáveis e outros identificadores.**

1. Qual o propósito em deixar um indicador de desatualizado no cabeçalho de um índice? Em um ambiente de multiprogramação, este indicador poderia ser encontrado “setado” por um programa porque outro programa está em processo de reindexação. Como o primeiro programa deveria responder a esta situação?

**Podemos criar um programa que lê/escreve um grande número de arquivos com diferentes características (número de campos por registro, comprimento de campos). Quanto mais informações houver no header, menos o o programa precisa saber sobre a estrutura específica de um arquivo em particular.**

1. Discuta o problema que ocorre quando você inclui o registro abaixo em seu arquivo, assumindo que o índice por compositor usado é o mostrado na Figura 6.9. Como você poderia resolver o problema sem grandes alterações na estrutura do índice de chave secundária?

LON 1259 Fidelio Beethoven Maazel



**Tirar o Beethoven da chave secundária e colocar ele na lista.**

1. O que é uma lista invertida? Quando ela é útil? Como ela é mantida em memória secundária? Esquematize o conteúdo de um índice secundário organizado como lista invertida para um arquivo de dados hipotético.

**Uma estrutura de dados que mapeia termos às suas ocorrências em um documento ou conjunto de documentos, armazenados em um banco de dados. As referências às chaves primárias associadas a cada chave secundária são mantidas em um arquivo sequencial separado, organizado segundo a entrada dos registros. É uma estratégia de indexação que permite a realização de buscas precisas e rápidas, em troca de maior dificuldade no ato de inserção e atualização de documentos.**

**Video1 Desenvolvimento, recuperação // Video2 recuperação, informação // Vídeo3 ontologia, semântica latente**

**Chave Primária // Chave Secundária**

**Desenvolvimento Video1**

**Recuperação Video1, Video2**

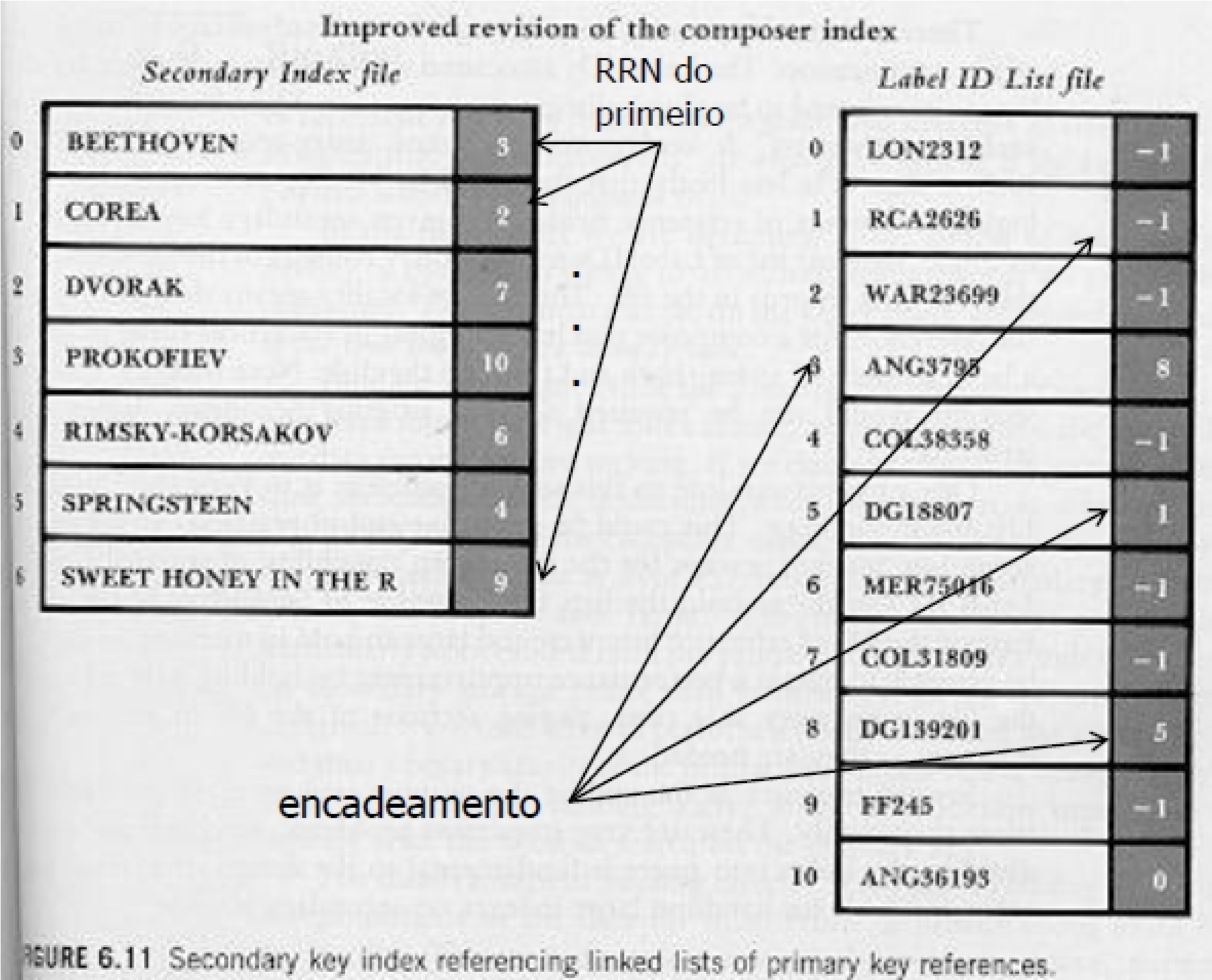
**Informação Video2**

**Ontologia Video3**

**Semântica latente Video3**

1. Como são alteradas as estruturas mostradas na figura 6.11 pela inclusão do registro abaixo?

LON 1259 Fidelio Beethoven Maazel



1. Suponha que um arquivo de dados com 6.000 registros, mantido em disco, deve ser ordenado em um computador cuja memória interna acomoda no máximo 600 registros por vez, usando o procedimento de intercalação. Considere que serão geradas 10 corridas de 600 registros cada, e que será realizada uma intercalação em 10 vias. Essa mesma área de memória interna é usada como buffer de entrada para leitura de dados do disco, e o sistema conta com um buffer de saída adicional que acomoda 200 registros.
   1. Durante a intercalação, quantos registros serão lidos de cada corrida cada vez que ela é acessada? Justifique.
   2. Quantos *seeks* serão realizados para ler dados durante o processo de intercalação

(excluindo a fase de geração de corridas)? Quantos serão realizados para escrever dados?

Justifique.