

Plano A: Evitando a duplicidade no uso de dados secundários

Contexto: É bastante comum o uso concomitante de dados primários (i.e., coletados na fonte) e secundários (i.e. literatura disponível) em pesquisas, principalmente naquelas onde o acesso aos dados primários são restritos ou se mostram insuficientes. Um exemplo disso é a coleta baseada em espécimes depositados em coleções científicas. Usualmente cada espécime possui um número tombo, que o identifica. Assim, quando a literatura disponibiliza os números tombos dos animais consultados, é possível montar um bancos de dados, baseado nos números tombo. Contudo, para evitar a duplicidade de dados, seria necessário acessar este banco de dados secundários toda vez que for coletar dados primários. A função que proponho realizará uma busca em uma base de dados secundários, pelos números tombos, considerando apenas a sigla da coleção (ex. NMNH) e suas eventuais variáveis (cobrindo assim erros de digitação).

Entrada de dados: Um arquivo .csv com todas as informações coletadas em literatura (espécie, localidade, sexo, ..., tombo) pelo usuário, separado por ponto e vírgula, e sem acentuação. A função terá os seguintes argumentos: dados (o arquivo .csv), x (sigla a ser buscada), exato (se a sigla a ser buscada será exata ou não), variação (variações possíveis do termo de busca), coluna.omit (omitir ou não colunas vazias nas linhas do termo x e suas variações).

Verificando os parâmetros de entrada:

1. data foi lido corretamente? Se não, escreve: "data não lido corretamente e/ou sem tabulação correta".
2. A sigla x a ser buscada existe em data? Se não, escreve: "argumento x não encontrado em data".
3. Todas as colunas vazias foram omitidas corretamente? Se não, escreve: "há colunas com dados que foram omitidas".

Saída de dados: Um arquivo .txt com uma tabela contendo as linhas onde a sigla x e suas variações (argumentos x e variação, respectivamente) aparecem no arquivo de entrada, e nas colunas, as variáveis já coletadas literatura para tais argumentos.

Pseudo-código:

1. Cria a função `secun` a partir da leitura de `data` e dos argumentos `x`, `exato`, `variacao`, `coluna.omit`.
2. Pesquisa apenas o termo `x`, se `exato=T`.
3. Pesquisa o termo `x` e suas variações, se `exato=F` e `variacao("termox1", "termox2", ..., "termox10")`.
4. Se `coluna.omit=T`, inicia um ciclo `for`, de `l` até o número de colunas do `data`, omitindo colunas vazias (apenas das linhas selecionadas pelo termo `x` e `variacao`).
5. Se `coluna.omit=F`, retorna a linha inteira onde aparece o termo `x` e aqueles adicionados em `variacao`.
6. Cria o objeto `tabela` onde são guardados as linhas e colunas selecionadas a partir dos argumentos da função.
7. Exporta o objeto `tabela` como um arquivo .txt a ser salvo no diretório de trabalho.

From:
<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:
http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2018:alunos:trabalho_final:silara_fatima:fun1 

Last update: **2020/08/12 06:04**