

|  |  |
| --- | --- |
| **ID:** | **UC03** |
| **Objetivo:** | Extrair dados de interesse da fonte “Scopus”. |
| **Requisitos:** | Estar conectado à internet na rede da Universidade Federal Fluminense (ou em outra rede que forneça acesso ao site scopus.com e à sua API).  Ter uma chave de acesso válida à API do “Scopus”. |
| **Atores:** | Usuário. |
| **Pré-condições:** | Os pacotes, linguagens, *frameworks* e bibliotecas usados pela ferramenta precisam estar instalados no ambiente.  A busca a ser realizada precisa estar definida no arquivo (JSON) de configurações/entradas da fonte.  A chave de acesso à API do “Scopus” precisa estar definida no arquivo (JSON) de configurações/entradas da fonte. |
| **Pós-condições:** | As informações obtidas deverão estar salvas em arquivo(s) no disco. |
| **Fluxo Principal:** | **Scrap Scopus**  1. O ator executa a ferramenta, através de linha de comando, com os parâmetros que definem a fonte “Scopus”;  2. A ferramenta busca os dados de interesse (definidos em um arquivo configurável do tipo JSON);  3. A ferramenta gera arquivo(s) de saída (conforme os parâmetros informados) no disco com as informações encontradas. |
| **Fluxos Alternativos:** | Não há. |
| **Erros/Exceções:** | O tratamento de erros (como requisições inválidas e *timeouts*) é feito pelo *framework* utilizado (mensagens de erro podem ser exibidas para o usuário).  Em caso de falha(s) na execução as pós-condições não serão verdadeiras (os dados de interesse não serão obtidos). |
| **Mensagens:** | *Logs* de saída do *framework* utilizado. |
| **Regras de negócio:** | **[RN07]** Apenas artigos publicados no últimos 10 (dez) anos devem ser considerados.  **[RN08]** As palavras-chave buscadas devem aparecer no título e/ou no *abstract* do artigo para que o mesmo seja considerado.  **[RN09]** Cada artigo deve ter uma fonte associada e cada fonte deve ter uma pontuação (*score*) que determina seu fator de impacto – apenas fontes que tenham a média dos últimos 3 (três) anos da sua respectiva pontuação maior que 1 (um) devem ser consideradas.  **[RN10]** Cada artigo deve ter um código DOI único associado – este código deve ser utilizado para evitar a coleta de artigos duplicados.  **[RN11]** O endereço de internet “https://api.elsevier.com/content/search/scopus" deve aceitar requisições do tipo GET com os parâmetros “query” e “count” com os valores das variáveis “termo buscado” (ou consulta – *query* – definida conforme os padrões da fonte “Scopus”) e “número desejado de resultados”, respectivamente, e deve ser utilizado para buscar os dados de interesse.  **[RN12]** A chave de acesso à API do “Scopus” deve estar contida no cabeçalho das requisições de busca (**RN11**), no campo de nome “X-ELS-APIKey”.  **[RN13]** O número de resultados obtidos pode ser limitado pelo usuário através das definições no arquivo de configurações/entradas da fonte.  **[RN14]** Os resultados das requisições de busca (**RN11**) devem estar no formato JSON e os registros encontrados devem estar na chave de nome “search-results” e conter, para cada artigo retornado pela busca, os campos “source-id”, “doi” e “url” com o ID da fonte, o código DOI e o endereço de internet associados, respectivamente.  **[RN15]** Os endereços de internet referentes aos artigos encontrados (obtidos pelas requisições definidas na **RN11** e conforme descrito na **RN14**) devem aceitar requisições do tipo GET com os parâmetros “httpAccept” e “fields” com os respectivos valores “application/json” e “description” e devem ser utilizados para coleta dos dados de interesse.  **[RN16]** Os resultados das requisições de acesso aos artigos (**RN15**) devem estar no formato JSON contendo, dentro da chave aninhada “abstracts-retrieval-response → item → bibrecord → head”, o *abstract* e o endereço da fonte do artigo nas chaves “abstracts” e “source → website”, respectivamente.  **[RN17]** O endereço dinâmico de internet “https://doi.org/{doi}” (onde “{doi}” deve conter a código DOI único de cada artigo obtido pelas requisições definidas na **RN11** e conforme descrito na **RN14**) deve aceitar requisições do tipo GET e deve ser utilizado para verificar a validade de cada artigo.  **[RN18]** O endereço dinâmico de internet “https://www.scopus.com/source/citescore/{id\_da\_fonte}.uri” (onde “{id\_da\_fonte}” deve conter o valor do campo “source-id” obtido pelas requisições definidas na **RN11** e conforme descrito na **RN14**) deve aceitar requisições do tipo GET e deve ser utilizado para obter os dados de pontuação (*score*) das fontes dos artigos encontrados.  **[RN19]** Os resultados das requisições definidas na **RN18** devem estar no formato JSON e conter o campo “rp”, definido na chave aninhada “yearInfo → {ano} → metricType”, onde {ano} assume o valor do ano desejado.  **[RN20]** As requisições definidas na **RN19** só devem ser consideradas caso a resposta contenha o campo ““documentType” com o valor “all”. |