**Usando o Vocabulary Extractor for C/C++**

O Vocabulary Extractor para C ou C++ deve ser executado a partir de um terminal usando a seguinte estrutura:

java -jar VocabularyExtractorForC.jar [-n project\_name] [-r revision] [-d project\_path] [-loc h i p] [-vxl result\_file\_name] [-csv result\_file\_name] [-func] [-file C H] [-msr]

java -jar VocabularyExtractorForCPP.jar [-n project\_name] [-r revision] [-d project\_path] [-loc h i p] [-vxl result\_file\_name] [-csv result\_file\_name] [-func] [-file C H] [-msr]

**Exemplo:** java -jar VocabularyExtractorForC.jar -n GenericProject -r 2.0.1 -d ./ProjetoGenerico -loc hi -vxl ./ProjetoGenerico/GenericProject.vxl -csv ./ProjetoGenerico/GenericProject.csv -func -file CH - msr

A tabela abaixo mantém o significado de cada parâmetro de execução do extrator.

Tabela 1. Parâmetros específicos do VocabularyExtractor para contagem de linhas de código

|  |  |
| --- | --- |
| **Parâmetro** | **Função** |
| -n *project\_name* | Identificador, um nome, dado ao projeto de cujo vocabulário será extraído. Essa informação será armazenada no arquivo VXL gerado; |
| -r *revision* | Identificador para a revisão do projeto de cujo vocabulário será extraído. Essa informação será armazenada no arquivo VXL gerado; |
| -d *project\_path* | Indica o path do código fonte do projeto de cujo vocabulário deverá ser extraído; |
| -loc | Ativa a contagem de loc. Uma das seguintes opções devem ser configuradas:  h - Mostra a quantidade de linhas ocupadas por headers (*diretivas de pré-processador);*  *i  -*Mostra a contagem de locs de funções internas;  *p -*Contagem de linhas físicas por arquivo C. Quando ativada a contagem não descreve entidades, mas arquivos de código fonte. Logo desconsidera as opções anteriores h ou i; |
| -vxl *result\_file\_name* | path e nome do arquivo onde o arquivo VXL resultante do processamento será gerado; |
| -csv  *result\_file\_name* | path e nome do arquivo onde o arquivo CSV resultante do processamento será gerado; |
| -func | Ativa a recuperação de informações internas as funções. Adiciona ao vocabulário das entidades os identificadores das chamadas de funções; |
| -file | Ativa a escolha de tipos de arquivos que devem ser processados, as seguintes opções devem ser escolhidas:  C: Ativa a extração de vocabulários para arquivos com a extensão .c ou no caso do Vocabulary Extractor For C++, arquivos .cpp.  H: Ativa a extração de vocabulários para arquivos de cabeçalho com a extensão .h. |
| - msr | Ativa a mensuração da quantidade de memória RAM utilizada durante a execução do VocabularyExtractor. |

Fonte: <http://www.softwarevocabulary.org>

**Usando o Terms Counter For C/C++**

O Terms Counter para C ou C++ deve ser executado a partir de um terminal usando a seguinte estrutura:

java -jar TermsCounter.jar -prop termsCounter\_file.properties -vxl input\_file\_name.vxl -csv csv\_result\_file.csv -txt txt\_result\_file.txt

**Exemplo:** java -jar TermsCounter.jar -prop ./files/termsCounter.properties -vxl ./files/termsCounterTest.vxl -csv ./files/teste.csv -txt ./files/teste.txt

A tabela abaixo descreve cada um dos parâmetros para execução do Terms Counter For C.

Tabela 2. Parâmetros específicos do TermsCounter para processar um arquivo de vocabulário VXL da linguagem C.

|  |  |
| --- | --- |
| **Parâmetro** | **Função** |
| -prop *termsCounter\_file*  *.properties* | path e nome do arquivo de propriedades onde estão configurados os parâmetros para os componentes VXLReader, IdentifierFilter e IR.  # ------ VXL Reader Properties ------  containerType=file  includeFunction=true  includeEnum=true  includeUnion=true  includeStruct=true  includeParameter=true  includeLocalVar=true  includeGlobalVar=true  includeFuncCall=true  includePrototype=true  includeLiterals=true  includeField=true  includeConstant=true  includePragma=true  includeDirective=true  includeError=true  includeMacro=true  includeComment=true  # ------ Identifier Filter Properties ------  english=yes  underscore=yes  camelcase=yes  intt=no  stopwords=yes  stemming=yes  orengo=no  convertToLowerCase=yes  limitTermLength=0  # ------ Information Retrieval Properties ------  irFunctionType=IR  tfVariant=ABSOLUTE  scoreCalculator=TF  distanceFunction=CANBERRA  # ------ Dispersion Measures Properties  dispersionMeasures=no  thresholdHE=10E-6  thresholdHC=10E-6  # ---------- Terms Counter Properties ----------  createMarginals=no  generateCSV=yes  generateTXT=yes |
| -vxl *input\_file\_name.vxl* | path e nome do arquivo VXL de entrada (vocabulário) que será processado pelo TermsCounter; |
| *-csv csv\_result\_file.csv* | path e nome do arquivo CSV resultante. Este contém em formato csv a Matriz de Medidas de IR (TF, IDF, TF-IDF) para cada Termo Distinto do projeto *em cada* Entidades do projeto; |
| *-txt txt\_result\_file.txt* | path e nome do arquivo TXT resultante. Contém informações sobre o vocabulário processado: Matriz n-Termos *por m-*Entidades de medida de IR (TF, IDF, TF-IDF), a medida de IR por termo em todo projeto, e a quantidade de entidades, termos distintos e termos totais para todo o projeto. |

Fonte: <http://www.softwarevocabulary.org>

A tabela abaixo descreve cada um dos parâmetros para execução do Terms Counter For C++.

Tabela 3. Parâmetros específicos do TermsCounter para processar um arquivo de vocabulário VXL da linguagem CPP.

|  |  |
| --- | --- |
| **Parâmetro** | **Função** |
| -prop *termsCounter\_file*  *.properties* | path e nome do arquivo de propriedades onde estão configurados os parâmetros para os componentes VXLReader, IdentifierFilter e IR.  # ------ VXL Reader Properties ------  containerType=file  includeFunction=true  includeClass=true  includeMethod=true  includeEnum=true  includeUnion=true  includeStruct=true  includeParameter=true  includeLocalVar=true  includeGlobalVar=true  includeFuncCall=true  includePrototype=true  includeLiterals=true  includeField=true  includeConstant=true  includePragma=true  includeDirective=true  includeError=true  includeMacro=true  includeComment=true  # ------ Identifier Filter Properties ------  english=yes  underscore=yes  camelcase=yes  intt=no  stopwords=yes  stemming=yes  orengo=no  convertToLowerCase=yes  limitTermLength=0  # ------ Information Retrieval Properties ------  irFunctionType=IR  tfVariant=ABSOLUTE  scoreCalculator=TF  distanceFunction=CANBERRA  # ------ Dispersion Measures Properties  dispersionMeasures=no  thresholdHE=10E-6  thresholdHC=10E-6  # ---------- Terms Counter Properties ----------  createMarginals=no  generateCSV=yes  generateTXT=yes |
| -vxl *input\_file\_name.vxl* | path e nome do arquivo VXL de entrada (vocabulário) que será processado pelo TermsCounter; |
| *-csv csv\_result\_file.csv* | path e nome do arquivo CSV resultante. Este contém em formato csv a Matriz de Medidas de IR (TF, IDF, TF-IDF) para cada Termo Distinto do projeto *em cada* Entidades do projeto; |
| *-txt txt\_result\_file.txt* | path e nome do arquivo TXT resultante. Contém informações sobre o vocabulário processado: Matriz n-Termos *por m-*Entidades de medida de IR (TF, IDF, TF-IDF), a medida de IR por termo em todo projeto, e a quantidade de entidades, termos distintos e termos totais para todo o projeto. |

Fonte: <http://www.softwarevocabulary.org>