
Universidade Federal de São Carlos - UFSCar
Departamento de Computação - DC
Programa de Pós-Graduação em Ciência da Computação - PPGCC

Módulo Perl para Extração de Citação em arquivos EndNote



Aluna: Juliana Lilian Duque
Orientador: Dr. Ricardo Rodrigues Ciferri
Área: Banco de Dados

EndNote

- Ferramenta comercial desenvolvido por Thomson Reuters;
 - Permite armazenar e organizar as referências encontradas nas buscas em bases de dados, possibilitando a inclusão automática de citações e referências;
 - Pode exportar bibliotecas de citação como texto simples, RTF, HTML ou XML;
 - Opcionalmente pode ser instalado de forma que sua funcionalidade aparece no Microsoft Word para citar quando você escreve, e outras funções;
 - Simplifica o trabalho.
-

■ Exemplos de arquivos .enw

%0 Conference Proceedings

%A Pablo Freire Matos

%A Ricardo Rodrigues Ciferri

%A Thiago Alexandre Salgueiro Pardo

%D 2009

%T Metodologia de pré-processamento textual para extração de informação em artigos científicos do domínio biomédico

%B WORKSHOP DE TESES E DISSERTAÇÕES EM BANCOS DE DADOS

%C Fortaleza, Ceará

%I Simpósio Brasileiro de Banco de Dados

%7 VIII

%P 7-12

%U <http://sca.dc.ufscar.br/download/files/Pablo.SBBD.WTDBD.2009.pdf>

%X O objetivo deste trabalho de pesquisa em nível de mestrado é extrair informação em artigos científicos

■ Objetivo

- ❑ Extrair todos os campos do arquivo EndNote;
- ❑ Extrair autores e título;
- ❑ Extrair título sem { , } , (,)
- ❑ Extrair primeiro e último autor;
- ❑ Inverter a ordem do autor (nome e sobrenome);
- ❑ Gerar XML para exportar os dados.

■ Metodologia

- ❑ Linguagem Perl
 - ❑ Sistema Operacional FreeBSD (UNIX)
 - ❑ Linguagem XML
-

Text::Refer

- Módulo Perl para manipulação de arquivos .enw
 - Biblioteca gratuita
 - Versão: 1.106
 - Data: 30 Abril de 1997

 - Download:
<http://search.cpan.org/CPAN/authors/id/E/ER/ERYQ/Text-Refer-1.106.tar.gz>
-

Text::Refer

Instalação da biblioteca no Unix

```
fetch -v http://search.cpan.org/CPAN/authors/id/E/ER/ERYQ/Text-Refer-1.106.tar.gz
```

```
tar zxvf Text-Refer-1.106.tar.gz
```

```
cd Text-Refer-1.106/
```

```
perl Makefile.PL
```

```
make
```

```
make test
```

```
make install
```

Instalação da interface para trabalhar com XML no Perl:

```
XML::Simple
```

```
shell> perl -MCPAN -e shell
```

```
cpan> install XML::Simple
```

Demonstração do código

```
#!/usr/local/bin/perl

use Text::Refer;
use strict;
use warnings;
use Data::Dumper;
use XML::Simple qw(XMLout);
my $arquivoxml = "texto.xml";
open (my $out, ">", "$arquivoxml") || die "erro no arquivo$\n";
```

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
```

```
- <EndNote>
  <type>Report</type>
  <city>São Carlos,</city>
  <date>2009</date>
  <publisher>Federal University of São Carlos</publisher>
  <page>16</page>
  <title>Technical Report "Sickle 'Cell' Anemia"</title>
  <title>Technical Report Sickle Cell Anemia</title>
  <firstauthor>Ana Pinto</firstauthor>
  <lastauthor>Marina Vieira</lastauthor>
</EndNote>
```

Bibliografia

- CPAN, Text::Refer. Disponível em <<http://search.cpan.org/~eryq/Text-Refer-1.106/lib/Text/Refer.pm>>. Acesso em 21 de Outubro de 2009.
 - Interjornal. “Segmento científico da Thomson Reuters lança EndNote X2 para Windows”. Disponível em <http://www.interjornal.com.br/noticia_pdf.kmf?noticia=7346389>. Acesso em 15 de Novembro de 2009.
 - EndNote. Disponível em <<http://www.endnote.com>>. Acesso em 15 de Novembro de 2009.
-