

RENA

Reconhecedor de Entidades

J.João Dias de Almeida
Edgar Alves

24 de Julho de 2006

- 1 Introdução
- 2 Configuração
 - Ontologia de classes
 - Tabela de contextos
 - Gazetteer de cultura geral
 - Sistema de Nomes
 - Tabela de conversão/adaptação de nomes
- 3 Alternativas HAREM
 - Rena: tripas
- 4 Conclusões

- RENA é:
 - Uma biblioteca Perl:
 - 1 extrair a lista das **Entidades** a partir de conjuntos de texto
 - 2 marcar **Entidades** num conjunto de texto
 - 3 baseada numa **configuração** alterável
 - Uma (ou mais?) scripts para fazer processamento de entidades
- Contexto = projecto IKF (Information + Knowledge + Fusion)
 - extracção de informação (rica)
 - classificação fuzzy e parcial
 - fusão (join) da informação extraída
 - *Ficheiro * + BaseEnt → BaseDoc + BaseEnt + ...*
 - Navegadores sobre caixas de mail, *papelada*

Configuração do RENA

Configuração de base

- 1 Tabela de contextos
- 2 Gazetteer de cultura geral
- 3 Sistema de Nomes
- 4 Tabela de conversão/adaptação de nomes

Planeando-se ainda...

- Tabela de contextos atributivos

```
X , no norte de Y  
o francês Z  
a atleta portuguesa A
```

Hierarquia de classes

- pessoa:
 - advogado
 - arquitecto
 - atleta:
 - futebolista
 - escritor:
 - poeta
 - jornalista
 - militar:
 - general
 - almirante
 - brigadeiro
 - sargento
 - tenente
 - capitão
 - musico:
 - compositor
 - pianista
 - político:
 - presidente da república
 - deputado

Tabela de contextos

```
cidade (de do da) => cidade !lctx
freguesia (de do da) => freguesia
distrito (de do da) => distrito
concelho (de do da) => concelho/90
estado (de do da) => estado
capital => cidade !lctx

(Rio Oceano Lago Mar Serra Cordilheira) => $_
Cabo (do de da) => cabo
Golfo (do de da) => golfo

(Lugar Largo Lg. Praça Rua R. Avenida) (de da do das dos)? => lugar
(Travessa Beco Quinta Viela Rotunda) (de da do das dos)? => lugar

#
# Monumentos
#

(Convento Mosteiro Igreja Ig. Palácio Museu Sé) (de da)? => monumento
```

...Vista sobre a projecção de uma ontologia

Rio Douro =

```
rio Douro
  IOF => rio
  AFLUENTES => rio Mau,
               rio Sousa,
               rio Varosa,
               rio Tâmega,
               rio Pinhão,
               ....
               rio Torto,
               rio Távora,
               rio Esla,
               rio Tua
  COMPRIMENTO => 927
  FOZ => Porto
  IN => Portugal,
       Espanha
  NASCE => serra do Urbião
  PASSA_EM => barragem do Pocinho,
              barragem de Miranda,
              barragem de Crestuma,
              Miranda do Douro,
              barragem do Carrapatelo,
              Régua,
```

Name-gazetteer

```
Diana      :nome: 35
Dias       :apelido: 252
Dina       :nome: 7
Dinis      :misto: [2 13 15]
Diogo      :misto: [8 2 6]
Domingues  :apelido: 48
Dora       :nome: 8
```

Tabela de conversão/adaptação de nomes

Alternativas Harem: sonhar não custa

- XML!
- $colec = (MetaData \times Texto)^*$
- taxonomia 2 níveis \rightarrow ontologia multi-nível de classes eventualmente...

```
país < entidade geográfica  
país < instituição administrativa
```

- $erro \neq dvida \neq vagueza$
- etiquetagem mais prática:
 - uma única tag Entidade
 - um atributo tipo $T = "pas"$
 - com possibilidade de alternativas? $T = "t1|t2"$
 - um atributo de unificação

$(\text{texto} \times \text{entidade})^* \rightarrow (\text{texto} \times \text{entidade})^*$

Conclusões e dívidas

- melhorar as regras de inferência e unificação e resumo
- criar um processador estrutural
- melhorar name-gazetteer estrangeiros
- MELHORAR!!!
- documentar melhor o interface de biblioteca Perl

```
forent($texto){ .... }
```