Utilização da Programação Declarativa para processamento do *CETEMPúblico*

Agostinho Monteiro
monteiro.agostinho@gmail.com
Júlio Barbas
julio.barbas@gmail.com
Nuno C. Marques
nmm@di.fct.unl.pt

PROPOR - Encontro 10 Anos Linguateca, 2008.

Tópicos

Introdução e Motivação

- Propõe-se um novo formato que possibilita a utilização directa da programação declarativa (e/ou funcional).

Introdução e Motivação

- Propõe-se um novo formato que possibilita a utilização directa da programação declarativa (e/ou funcional).
- Esta proposta deve ser integrada com as soluções e formatos já existentes.

Introdução e Motivação

- Propõe-se um novo formato que possibilita a utilização directa da programação declarativa (e/ou funcional).
- Esta proposta deve ser integrada com as soluções e formatos já existentes.
- Pretende potenciar a sua interoperabilidade por via de uma filosofia ***SOA.***

Introdução e Motivação

- Propõe-se um novo formato que possibilita a utilização directa da programação declarativa (e/ou funcional).
- Esta proposta deve ser integrada com as soluções e formatos já existentes.
- Pretende potenciar a sua interoperabilidade por via de uma filosofia ***SOA.***

+ Extremamente fácil descrever o conhecimento e efectuar inferências e deduções sobre este mesmo conhecimento (e.g. [Marques et al.(2007)]).

Introdução e Motivação

- Propõe-se um novo formato que possibilita a utilização directa da programação declarativa (e/ou funcional).
- Esta proposta deve ser integrada com as soluções e formatos já existentes.
- Pretende potenciar a sua interoperabilidade por via de uma filosofia ***SOA.***

+ Extremamente fácil descrever o conhecimento e efectuar inferências e deduções sobre este mesmo conhecimento (e.g. [Marques et al.(2007)]).
+ Utilizando gramáticas, expressas por ***DCGs***, podem-se extraír representações lógicas relativas a determinados padrões de etiquetas que ocorrem em textos previamente marcados. (e.g. [Pereira e Shieber(1987)]).

Codificação de Informação no Formato TXT/2

TXT/2 para *Presidente da República*

```
txt(1, [w([1,wd='Presidente da República', tag='PROP',
        cw=[w([1-1, wd='Presidente', bw=presidente, gen=masc,
num=sing]),
                w([1-2, wd='da',
                        cw=[w([1-2-1, wd='de', tag='PREP']),
                                w([1-2-2, wd='a', tag='DET', gen=fem])]]),
                w([1-3, wd='República', gen=fem])]
        ])] ).
```

Outros Formatos

- Árvores deitadas ***(Floresta)***
- ***XML***
- ***Tiger-XML***
- ***SimTreeML***

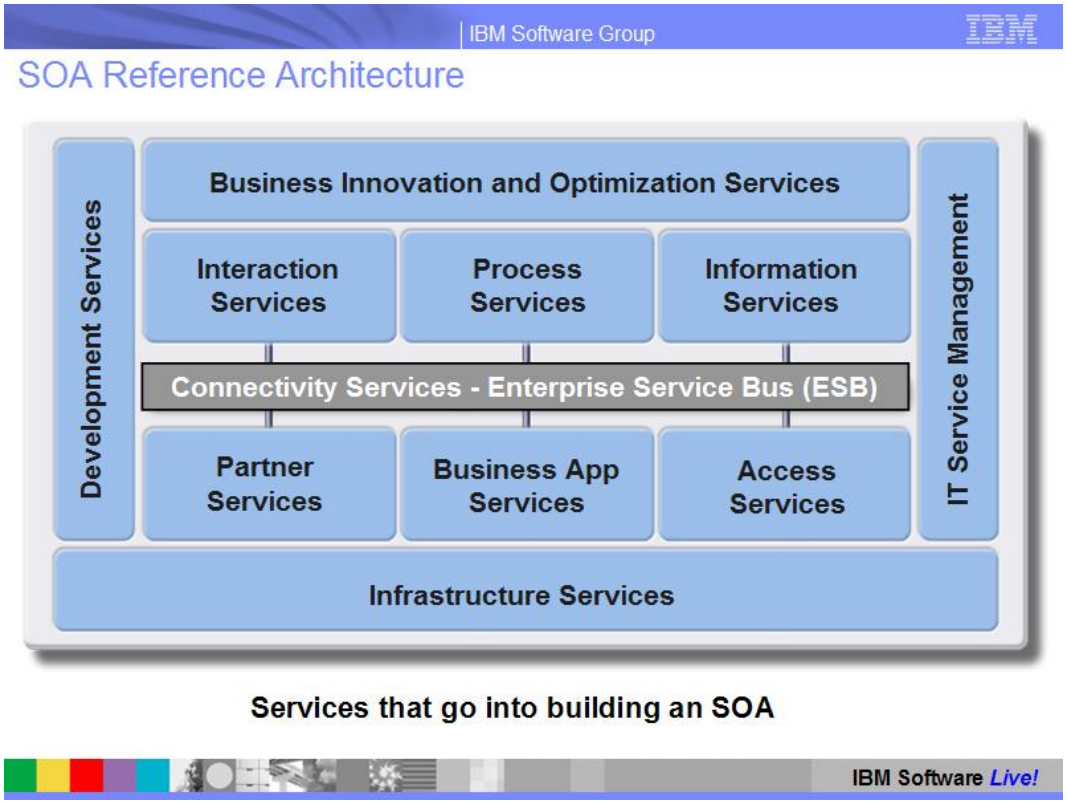
Inferencia de número em *Presidente da Republica*

Código ***Prolog***

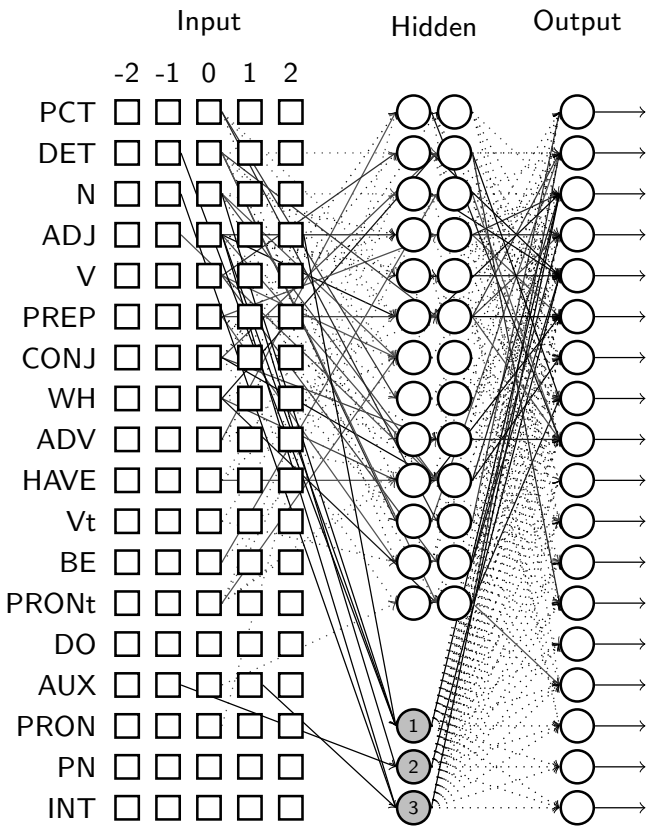
```
get_wd_f(W,F, IDW, TXT2) :-
        member(w( [IDW | WL]), TXT2),
        member(wd=W, WL),
        member(F, WL).
get_wd_f(W, F, IDW, TXT2) :-
        member(w([IDW | WL]), TXT2),
        member(cw=CW, WL),
        get_wd_f(W, F, _IDW, CW).
num(w(Wd), num=Num) :-
        member(cw=[H | T], Wd),
        get_wd_f(_W1, tag='PREP', _IDW1, T),
        get_wd_f(_W2, num=Num, _IDW2, [H]).
g :-
        txt(I,[W | _]),
        num(W,G).
```

Arquitectura Orientada aos Serviços

Utilizar (e reutilizar) a ferramenta adequada em cada momento, de forma transparente



Sistemas Neuro-Simbolicos [Marques et al.(2007)]



Regra
1) $n \leftarrow det_{-1}, adj_0, prep_1, adj_2$
the *urgency* of disarmament, the *sexes* in West
2) $v \leftarrow aux_{-1}, n_0$
would *force*, would *amount*
3) $n \leftarrow n_{-1}, adj_0, aux_1$
radar *screens* would register Soviet missiles

- Rede treinada com 10, 000 exemplos, após inserção de regras.
- As unidades a cinzento correspondem às 3 regras inseridas.
- A escala cinzenta codifica os pesos.
- Pesos negativos representados por linhas pontilhadas.
- Omitidas as ligações fracas.
- Indica a categoria mais provavel no lexico, nao a correcta.
- Regra 2: Se o dicionário diz: Existe um substantivo na posição actual (n_0) e um verbo auxiliar na posição -1 (aux_{-1}), então a palavra é mesmo um verbo (v).

Web Semântica

- Identificação de Padrões.

Web Semântica

- Identificação de Padrões.
- Localização, na Web, dos items pretendidos com redução/anulação de ambiguidade nas respostas obtidas.

Web Semântica

- Identificação de Padrões.
- Localização, na Web, dos items pretendidos com redução/anulação de ambiguidade nas respostas obtidas.
- Adicionar semântica aos dados (i.e. Inserção de contexto e ontologias).

Web Semântica

- Identificação de Padrões.
- Localização, na Web, dos items pretendidos com redução/anulação de ambiguidade nas respostas obtidas.
- Adicionar semântica aos dados (i.e. Inserção de contexto e ontologias).

+ Com XML sabemos como são os dados mas não o que são. Não há semântica.

Web Semântica

- Identificação de Padrões.
- Localização, na Web, dos items pretendidos com redução/anulação de ambiguidade nas respostas obtidas.
- Adicionar semântica aos dados (i.e. Inserção de contexto e ontologias).

+ Com XML sabemos como são os dados mas não o que são. Não há semântica.
+ Com o txt//2 dizemos o que são.

Web Semântica

- Identificação de Padrões.
- Localização, na Web, dos items pretendidos com redução/anulação de ambiguidade nas respostas obtidas.
- Adicionar semântica aos dados (i.e. Inserção de contexto e ontologias).

+ Com XML sabemos como são os dados mas não o que são. Não há semântica.
+ Com o txt//2 dizemos o que são.
+ Um programa Prolog pode mesmo alterar-se a si próprio - Metaprogramação.

Conclusões

Uma estrutura deste tipo permite:

- Contribuir para a utilização de módulos com maior poder dedutivo.
- Funcionar como um armazem de anotações efectuadas sobre o texto.
- Estimular a interoperabilidade entre diversas aplicações, promovendo uma filosofia ***SOA***.
- Potenciar a partilha de informação disponibilizada com o projecto da ***Linguateca***.

Trabalho futuro:

- Desenvolvimento de módulos extractores de representações lógicas, relativas a padrões de etiquetas que ocorrem em textos previamente marcados.
- Melhorar os resultados obtidos com o Sistema Neuro-Simbólico de etiquetagem Morfo-Sintáctica.

Referências (Recursos)


Linguateca - bosque sintáctico.
http://www.linguateca.pt/Floresta/.


TIGER PROJECT Linguistic Interpretation of a German Corpus.
http://www.ims.uni-stuttgart.de/projekte/TIGER/.


Geoffrey Sampson.
English for the Computer: The SUSANNE Corpus and Analytic Scheme, Oxford University Press, Oxford, 1995.
ISBN 0198240236.

Referências (Base)

Nuno Marques, Sebastian Bader, Vitor Rocio e Steffen Hölldobler.
Neuro-symbolic word tagging.
Em José Neves, Manuel Filipe Santos e José Machado, editores, *New Trends in Artificial Intelligence*. Associação Portuguesa para a Inteligência Artificial (APPIA), Guimarães. Portugal, Dezembro de 2007.
ISBN 13 978-989-9561.

Fernando C. N. Pereira e Stuart M. Shieber.
Prolog and Natural-Language Analysis — Digital Edition, Microtome Publishing, 1987.
http://www.mtome.com/Publications/PNLA/pnla.html

Jan Wielemaker.
Swi-prolog 5.6 reference manual.
http://gollem.science.uva.nl/cgi-bin/nph-download/SWI-Prolog/refman/refman.pdf

Fim

Questões?