ONTOLP: CONSTRUÇÃO SEMI-AUTOMÁTICA DE ONTOLOGIAS A PARTIR DE TEXTOS DA LÍNGUA PORTUGUESA

Mestrando: Luiz Carlos

Orientadora: Renata Vieira

Laboratório de Engenharia da Linguagem

SUMÁRIO

- Introdução
- Objetivo
- Metodologia
- OntoLP
- Avaliação
- Recursos
- Considerações Finais

INTRODUÇÃO (MOTIVAÇÃO)

- Web Semântica
  - "*Para o funcionamento da Web Semântica, computadores devem ter acesso a coleções estruturadas de informação e conjuntos de regras que eles possam usar para conduzir raciocínio automático*" (Berners-Lee, 2001)

- Construção Manual de Ontologias
  [Brewster et al.,2003]
  - Complexo
  - Tedioso
  - De alto custo

INTRODUÇÃO (OBJETIVO)

- Propor e avaliar técnicas para a construção automática de ontologias a partir de textos da língua portuguesa com base em técnicas já desenvolvidas para outras línguas

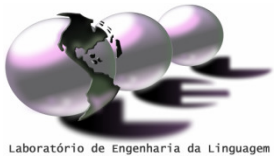
INTRODUÇÃO (CONSTRUÇÃO AUTOMÁTICA DE ONTOLOGIAS)

- Pontos Chave:
  - Etapas Executadas
    - Extração de Termos
    - Organização Hierárquica dos Termos
  - Fonte de Conhecimento
    - Textos

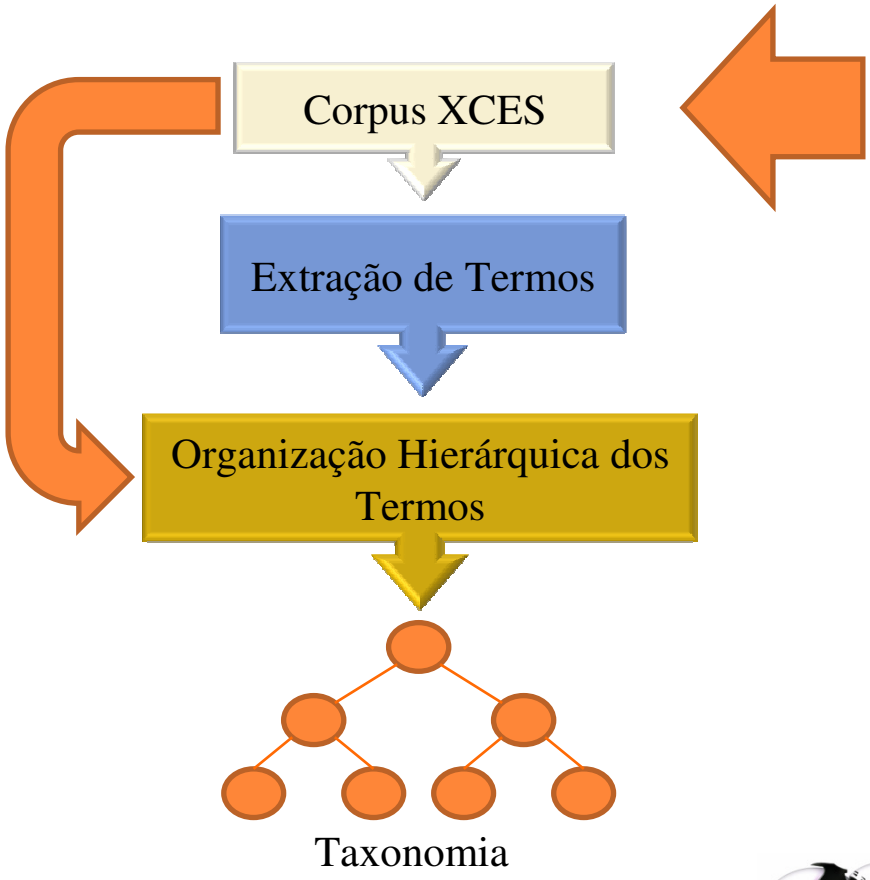
INTRODUÇÃO (CONSTRUÇÃO A PARTIR TEXTOS)

- Principais Métodos (Extração de Termos):
  - Estatísticos
  - Lingüísticos
  - Híbridos

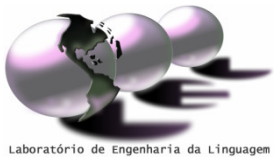
- Principais Métodos (Organização Hierárquica):
  - Baseado em Termos Complexos
  - Baseado em Padrões
  - Baseado nas medidas de Especificidade e Similaridade



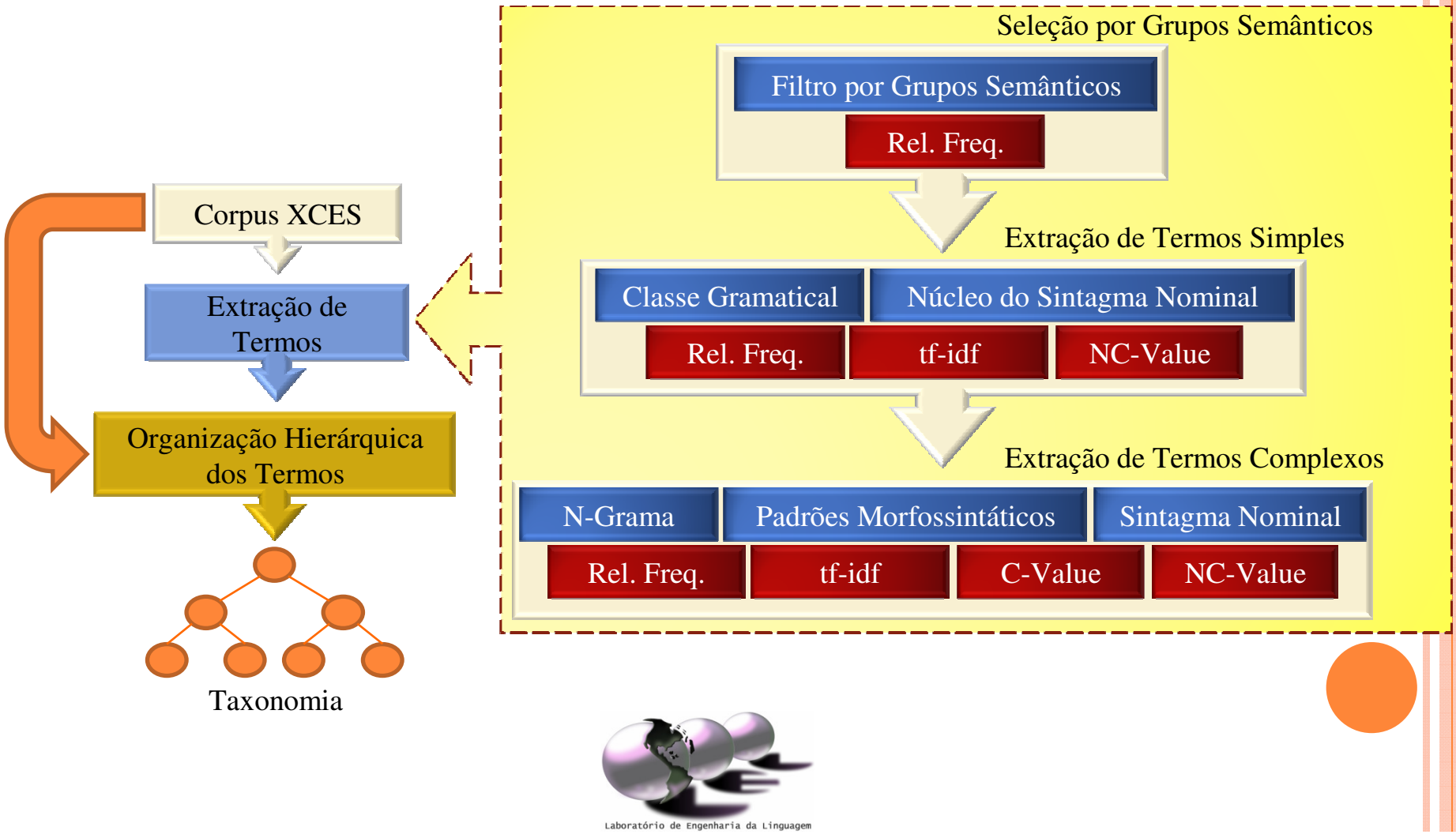
METODOLOGIA PROPOSTA



- Formato XCES:
  - Token
  - PoS
    - Morfológico
    - Semântico (tags semânticas)
  - Phrase
    - Sintático

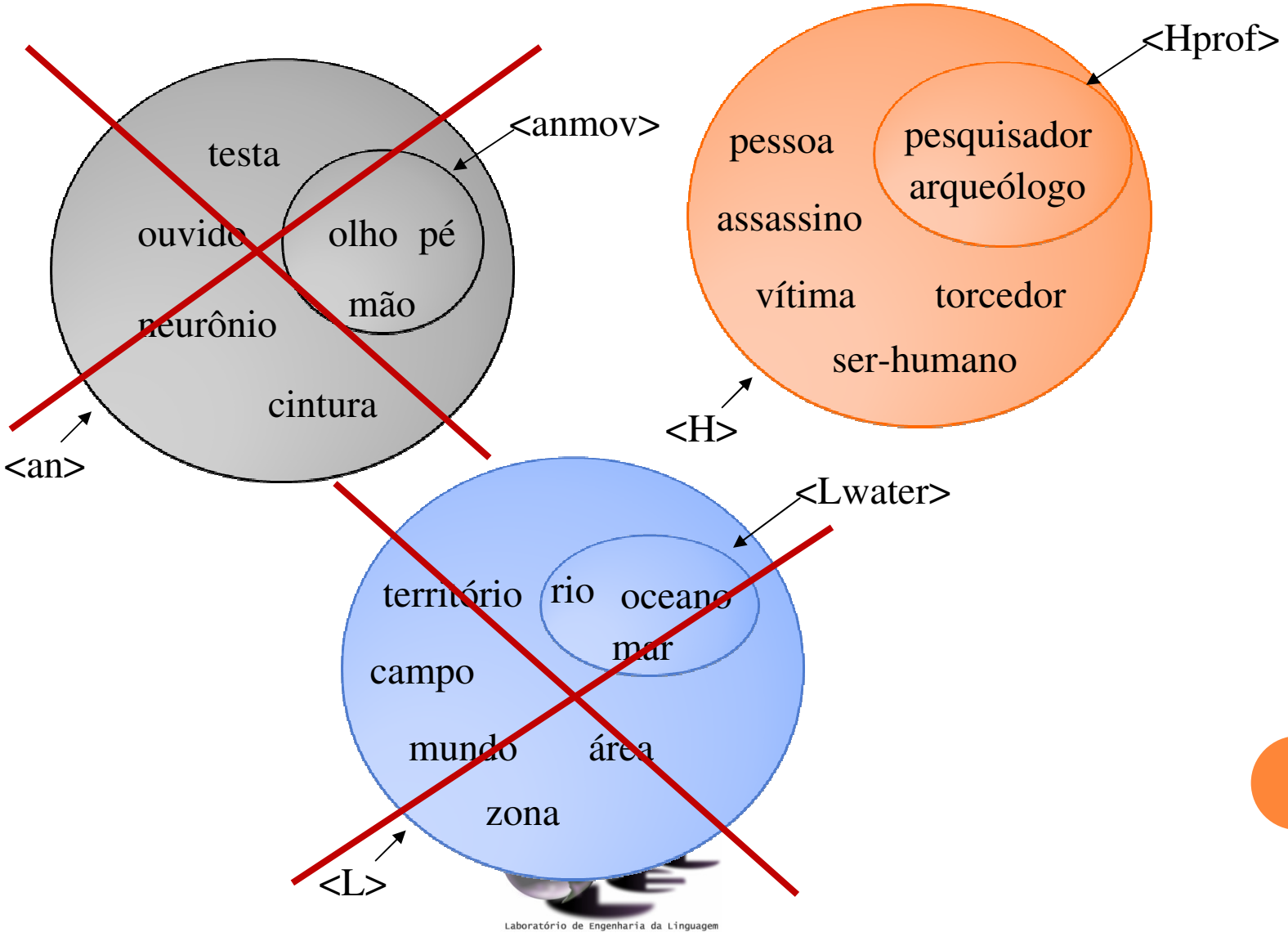


METODOLOGIA (EXTRAÇÃO DE TERMOS)

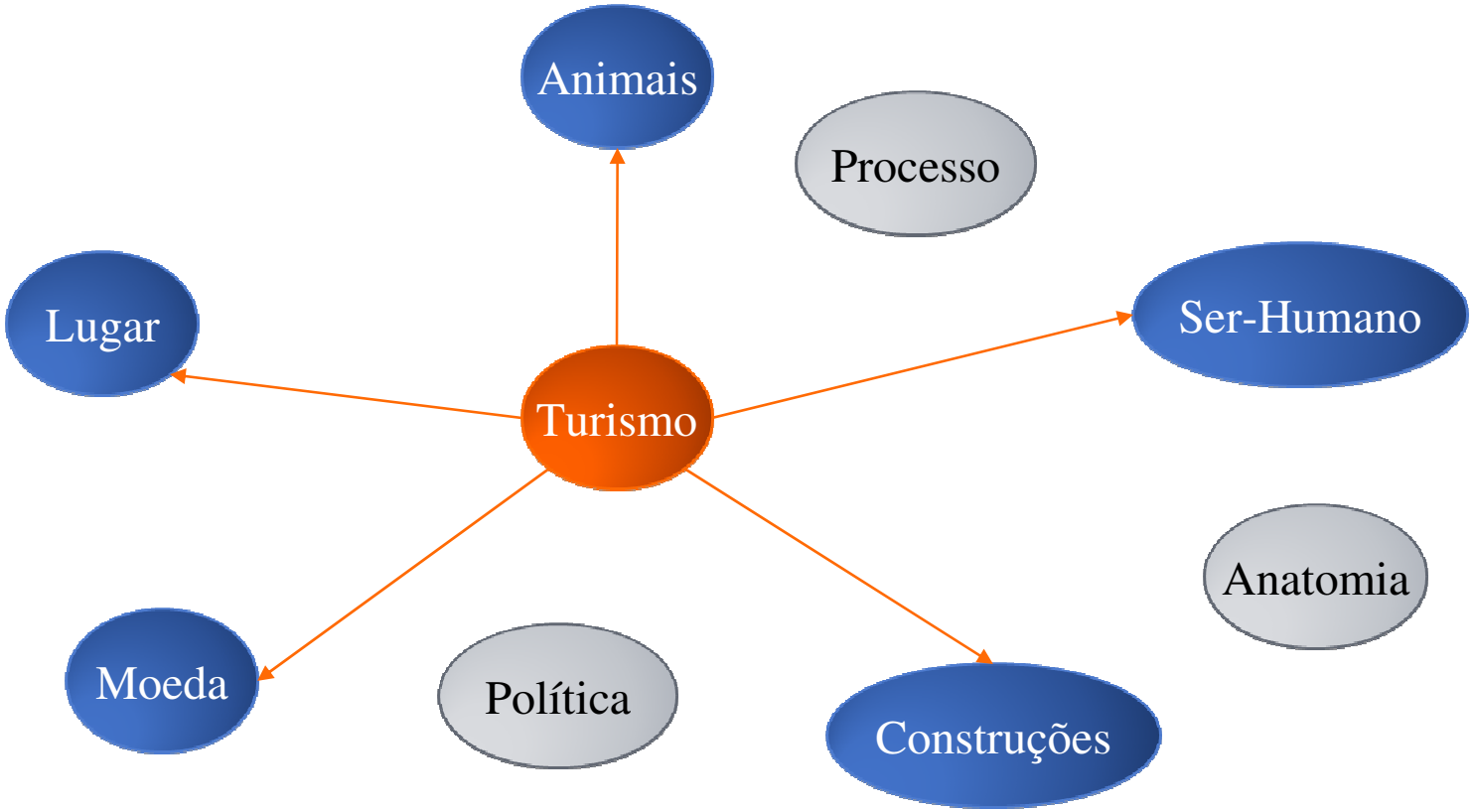


Laboratório de Engenharia da Linguagem

METODOLOGIA (GRUPOS SEMÂNTICOS)

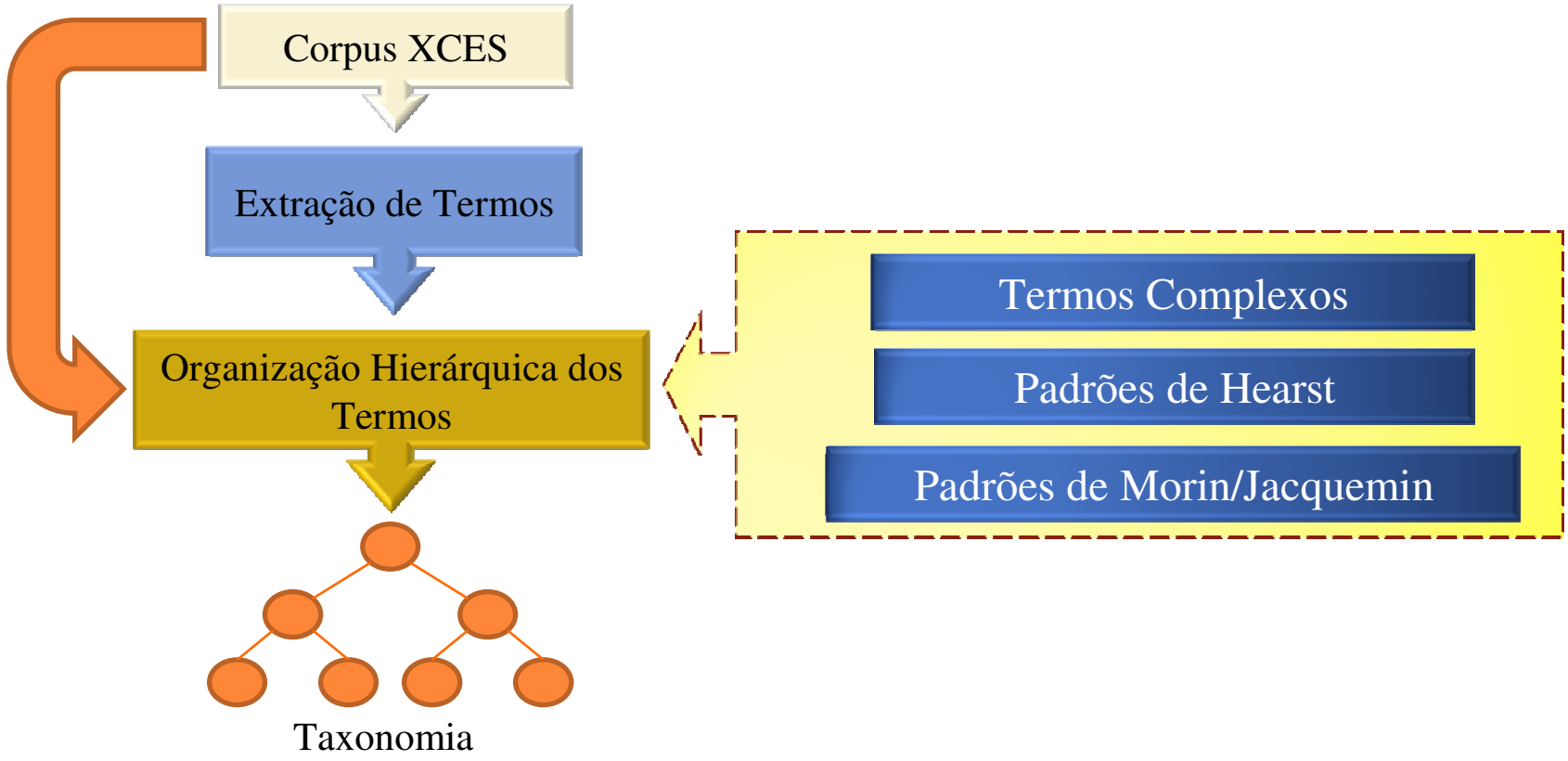


METODOLOGIA (GRUPOS SEMÂNTICOS)



Laboratório de Engenharia da Linguagem

METODOLOGIA (ORGANIZAÇÃO HIERÁRQUICA DOS TERMOS)

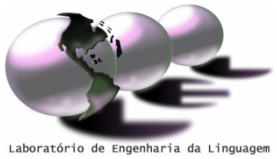


Laboratório de Engenharia da Linguagem

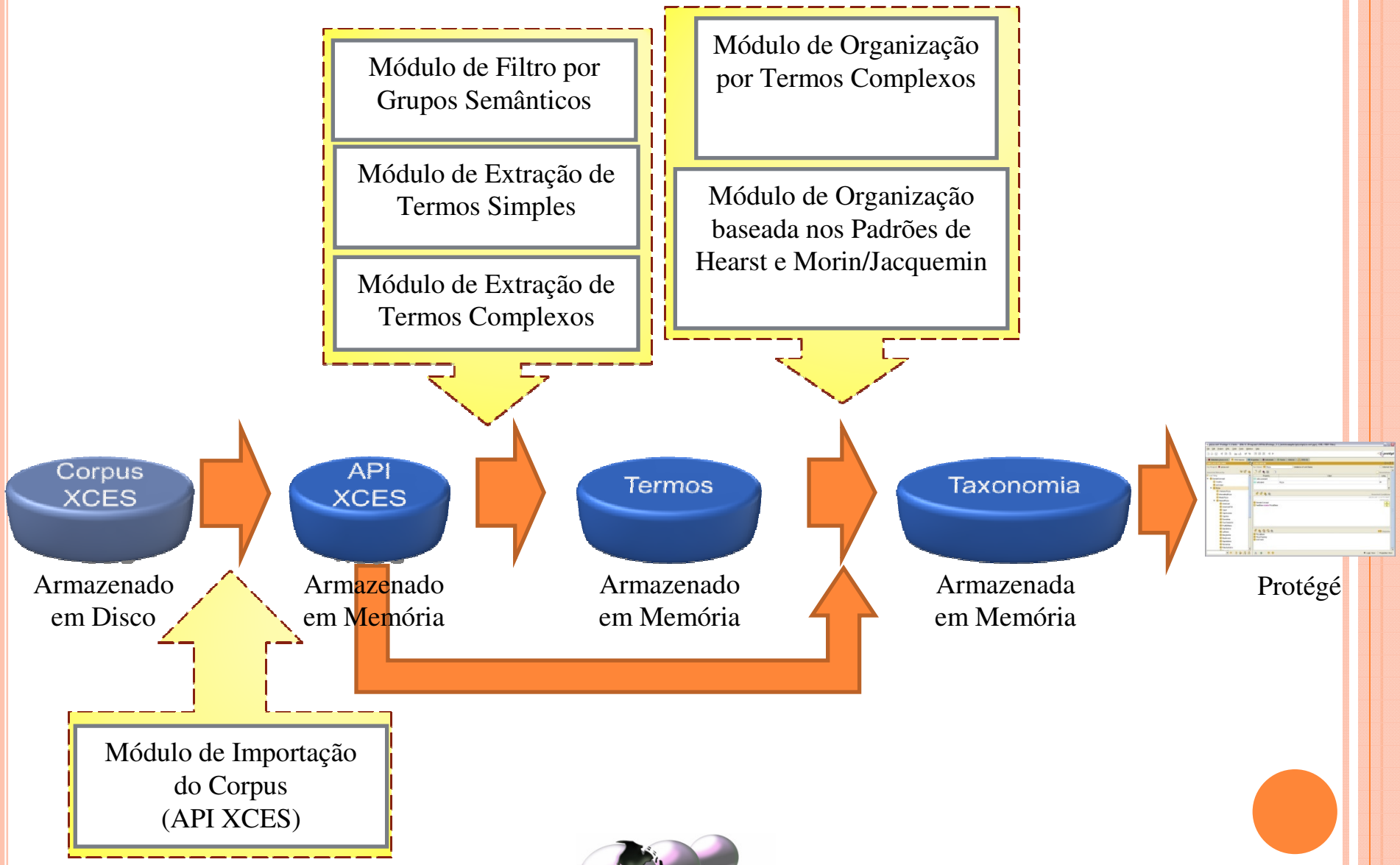
PADRÕES DE HEARST E MORIN/JACQUEMIN

Hearst/Baségio	Baségio/Termos Complexos
SUB como {(**SUB**,) * (oule)} **SUB**	**T** como {(**T**,) * (oule)} **T**
SUB tal(is) como {(**SUB**,) * (oule)} **SUB**	**T** tal(is) como {(**T**,) * (oule)} **T**
tal(is) **SUB** como {(**SUB**,) * (oule)} **SUB**	tal(is) **T** como {(**T**,) * (oule)} **T**

Morin/Baségio	Baségio/Termos Complexos
SUB1 (LIST_**SUB2**)	**T1** (LIST_**T2**)
SUB1: LIST_**SUB2**	**T1**: LIST_**T2**
SUB1 e (notadamente - em particular) **SUB2**	**T1** e (notadamente - em particular) **T2**

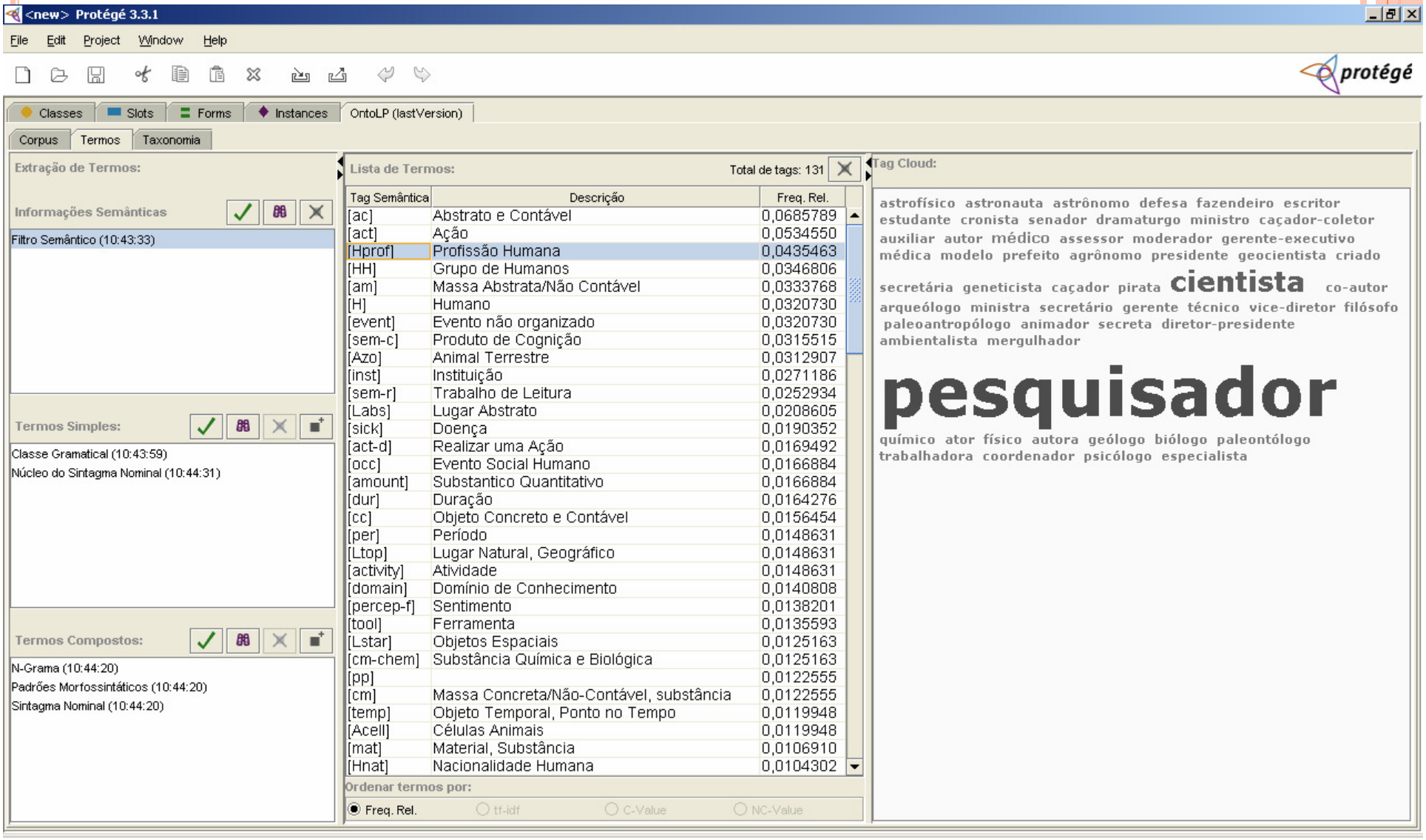


ONTOLP

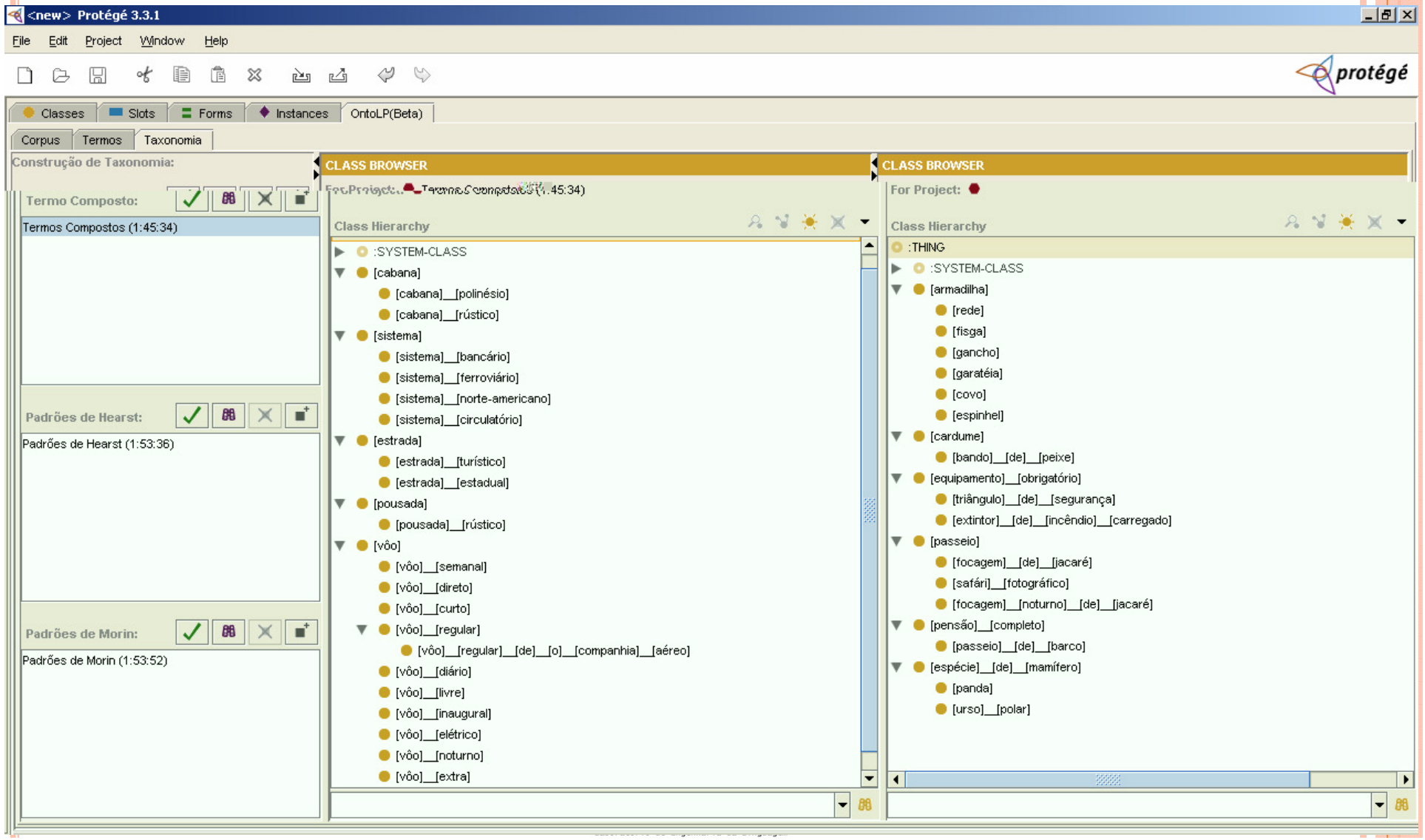


Laboratório de Engenharia da Linguagem

ONTOLP (INTERFACE DE EXTRAÇÃO DE TERMOS)



ONTOLP (INTERFACE DE ORGANIZAÇÃO HIERÁRQUICA)



AVALIAÇÃO

- Avaliação dos Métodos:
  - Precisão, Abrangência e *F-measure*
  - Recursos Necessários:
    - Corpus
    - Ontologia de Referência
      - Construída Manualmente
- Avaliação feita pelo usuário:
  - Avaliação das Funcionalidades do plug-in
  - Avaliação da etapa de Extração de Termos
    - Seleção por Grupos Semânticos

AVALIAÇÃO/RECURSOS

- Avaliação dos Métodos:
  - *CórpusEco* (Ecologia)
    - Lista de Termos uni, bi e trigrama
    - Ontologia do subdomínio "Ecologia das Comunidades"

- Avaliação feita pelo usuário:
  - *NanoTerm* (Nanotecnologia & Nanociência) (USP/SCar)
  - *JPED* (Pediatria) (UFRGS)

CONSIDERAÇÕES FINAIS

- Principal Dificuldade:
  - Escassez de recursos de avaliação

- Principais Contribuições:
  - Avaliação do uso de informações semânticas na construção de ontologias para o Português
  - Criação de uma ferramenta de uso geral para auxílio ao processo de construção de ontologias
  - Desenvolvimento dos módulos de avaliação automática das etapas executadas

CONSIDERAÇÕES FINAIS

- EDITAL CT-INFO 2007 - "Grandes Desafios da Computação"
  - Objetivo Geral:
    - Organizar recursos, métodos de construção e avaliação, e publicações da área de Ontologias para a Língua Portuguesa;
  - Objetivos Específicos:
    - Impulsionar as pesquisas na área;
    - Promover a interação entre grupos de pesquisa;
    - Disponibilizar subsídios para comparação entre métodos (*Benchmark*)