

ChatterEDU 2.0: ferramenta web de auxílio acadêmico utilizando chatterbot

Aluno: Eduardo Ferreira Gehrke

Orientadora: Joyce Martins

Roteiro

- Introdução
- Objetivos
- Fundamentação teórica
- Trabalhos correlatos
- Requisitos
- Especificação
- Implementação
- Demonstração da implementação
- Resultados e discussões
- Conclusões e sugestões

Introdução

- Interesse do autor por Processamento de Linguagem Natural (PLN);
 - Extensão do trabalho de Martins (2016).
- Importância da tecnologia na educação;
- Geração automática de questões de múltipla escolha:
 - Assunto pouco explorado na língua portuguesa.

Objetivos

- Objetivo geral: introduzir a geração de questões de múltipla escolha ao ChatterEDU 1.0;
- Objetivos específicos:
 - Buscar opções de respostas incorretas em páginas virtuais;
 - Migrar o *chatterbot* para a plataforma web;
 - Aperfeiçoar o tratamento de papéis semânticos já existentes.

Fundamentação Teórica

- ChatterEDU 1.0 (Martins, 2016):
 - *Chatterbot* desktop;
 - Desenvolvido em Java;
 - Gera base de conhecimento AIML, interpretada por meio do Program AB.

Fundamentação Teórica

- ChatterEDU 1.0 (Martins, 2016):
 - Texto de entrada é classificado pelo *parser* Palavras.

```
</β>  
Brasília [Brasília] <*> PROP F S @SUBJ> §TH #1->2  
é [ser] <fmc> <vK> V PR 3S IND VFIN @FS-STA §PRED #2->0  
a [o] DET F S @>N #3->4  
capital [capital] <Lciv> N F S @<SC §ATR #4->2  
de [de] <sam-> PRP @N< #5->4  
o [o] <-sam> DET M S @>N #6->7  
Brasil [Brasil] <*> PROP M S @P< §LOC #7->5  
.[.] PU @PU #8->0 </β> </β>
```

Fundamentação Teórica

- ChatterEDU 1.0 (Martins, 2016):
 - Com o resultado obtido, são geradas perguntas e, posteriormente, a base AIML.

```
<category>
  <pattern>^ qual ^ é ^ capital ^ Brasil ^</pattern>
  <template>
    <random>
      <li> Brasília </li>
      <li> Brasília é a capital do Brasil </li>
    </random>
  </template>
</category>
```

Fundamentação Teórica

- ChatterEDU 1.0 (Martins, 2016):
 - O usuário pode fazer perguntas, e solicitar que o robô as faça.

VISITANTE: pergunte algo sobre o texto
ChatterEdu: Qual é a capital do Brasil?

VISITANTE: Brasília
ChatterEdu: Acertou em cheio hein!

VISITANTE: qual é a capital do Brasil?
ChatterEdu: Brasília é a capital do Brasil

Fundamentação Teórica

- ChatterEDU 1.0 (Martins, 2016):
 - Porém, a geração de perguntas e respostas apresenta inconsistências:

O Brasil foi descoberto no mês de abril.

VISITANTE: pergunte algo sobre o texto
ChatterEdu: Quando O Brasil foi descoberto abril?

Fundamentação Teórica

- Geração automática de questões de múltipla escolha:
 - Geração de perguntas e respostas;
 - Geração de distratores.

Fundamentação Teórica

- Técnicas de geração automática de perguntas e respostas:
 - baseada em sintaxe;
 - baseada em semântica;
 - baseada em *template*.

Fundamentação Teórica

- Técnicas de geração automática de perguntas e respostas:
 - baseada em sintaxe:
 - Relações de concordância entre palavras em uma sentença.
 - Ambiguidade sintática: “O treinador disse ao auxiliar que sua estratégia seria eficaz.”

Fundamentação Teórica

- Técnicas de geração automática de perguntas e respostas:
 - baseada em semântica:
 - Sentido e interpretação das palavras;
 - “Richard Bird propôs a máquina NORMA.”

Fundamentação Teórica

- Técnicas de geração automática de perguntas e respostas:
 - baseada em semântica:
 - Sentido e interpretação das palavras;
 - “Richard Bird propôs a máquina NORMA.”
 - Pessoa: pronome interrogativo quem.

Fundamentação Teórica

- Técnicas de geração automática de perguntas e respostas:
 - baseada em *template*:
 - Substituição de espaços vazios em *templates* pré-definidos por palavras do texto de entrada.

Fundamentação Teórica

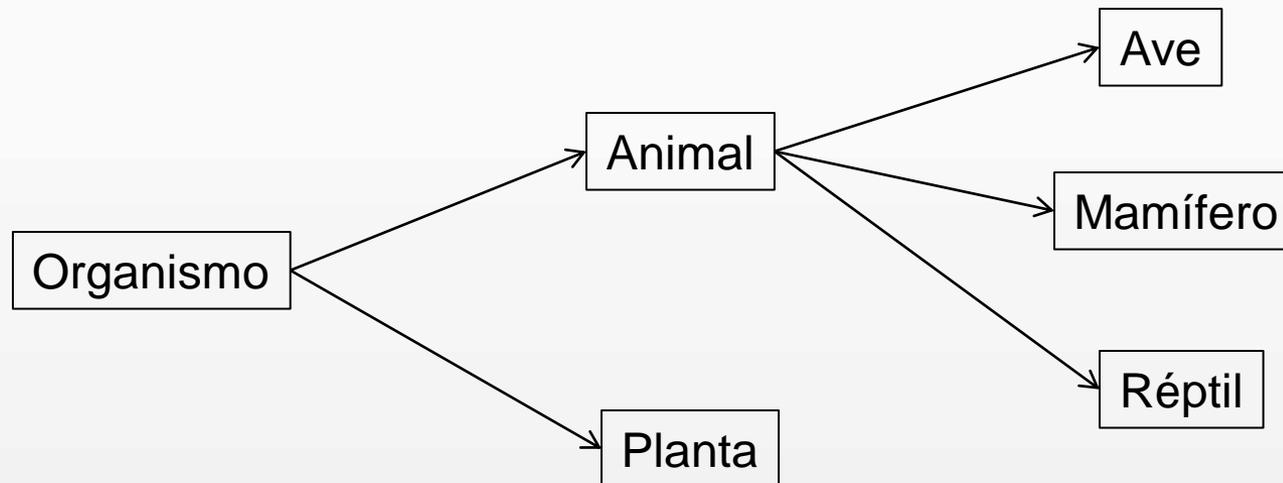
- Técnicas de geração automática de perguntas e respostas:
 - baseada em *template*:
 - Substituição de espaços vazios em *templates* pré-definidos por palavras do texto de entrada.
 - Quais são as características geográficas próprias de *Local*? (LABUTOV; BASU; VANDERWENDE, 2015).

Fundamentação Teórica

- Técnicas de geração automática de distratores:
 - uso de ontologias;
 - busca no próprio texto de entrada.

Fundamentação Teórica

- Técnicas de geração automática de distratores:
 - uso de ontologias:



Fundamentação Teórica

- Técnicas de geração automática de distratores:
 - busca no próprio texto de entrada:
 - O Brasil fica na América do Sul.
 - O Canadá fica na América do Norte.
 - A Alemanha fica na Europa.

Fundamentação Teórica

- Técnicas de geração automática de distratores:
 - busca no próprio texto de entrada:
 - O Brasil fica na América do Sul.
 - O Canadá fica na América do Norte.
 - A Alemanha fica na Europa.

Trabalhos Correlatos

- Brito (2017):
 - *Chatterbot* web baseado em uma FAQ:
 - Assunto: curso de Sistemas de Informação da Universidade Federal do Ceará.
 - Gera respostas às perguntas do usuário.

Trabalhos Correlatos

- Souza e Moraes (2015):
 - *Chatterbot* web baseado em uma FAQ:
 - Assunto: doenças.
 - Gera respostas às perguntas do usuário;
 - Faz uso das tecnologias VISL (Palavras), linguagem AIML e *parser* JSOUP.

Trabalhos Correlatos

- Curto (2010):
 - Gerador de perguntas e respostas, com a inclusão de distratores a partir do texto de entrada;
 - Perguntas são feitas ao usuário.

Requisitos funcionais

- RF01: possibilitar a interação com o usuário utilizando a língua portuguesa;

Requisitos funcionais

- RF02: permitir que o usuário forneça como entrada um texto gramaticalmente correto:
 - Ordem sujeito, verbo e objeto.

Requisitos funcionais

- RF03: gerar perguntas e respostas consistentes, evitando erros ortográficos ou de concordância verbal;

Requisitos funcionais

- RF04: impedir que a base de conhecimento gerada seja sobrescrita a cada execução;

Requisitos funcionais

- RF05: permitir a geração de questões de múltipla escolha;

Requisitos funcionais

- RF06: acessar e processar páginas web para obter elementos distratores.

Requisitos não funcionais

- RNF01: ser desenvolvido na linguagem Java, no ambiente de desenvolvimento NetBeans;

Requisitos não funcionais

- RNF02: estar disponível em uma plataforma web, por meio da ferramenta Apache Tomcat;

Requisitos não funcionais

- RNF03: realizar requisições ao site Palavras para efetuar a análise morfológica:
 - Obter o resultado para gerar a base AIML.

Requisitos não funcionais

- RNF04: extrair informações de páginas HTML por meio do *parser* JSOUP.

Especificação

Arquitetura

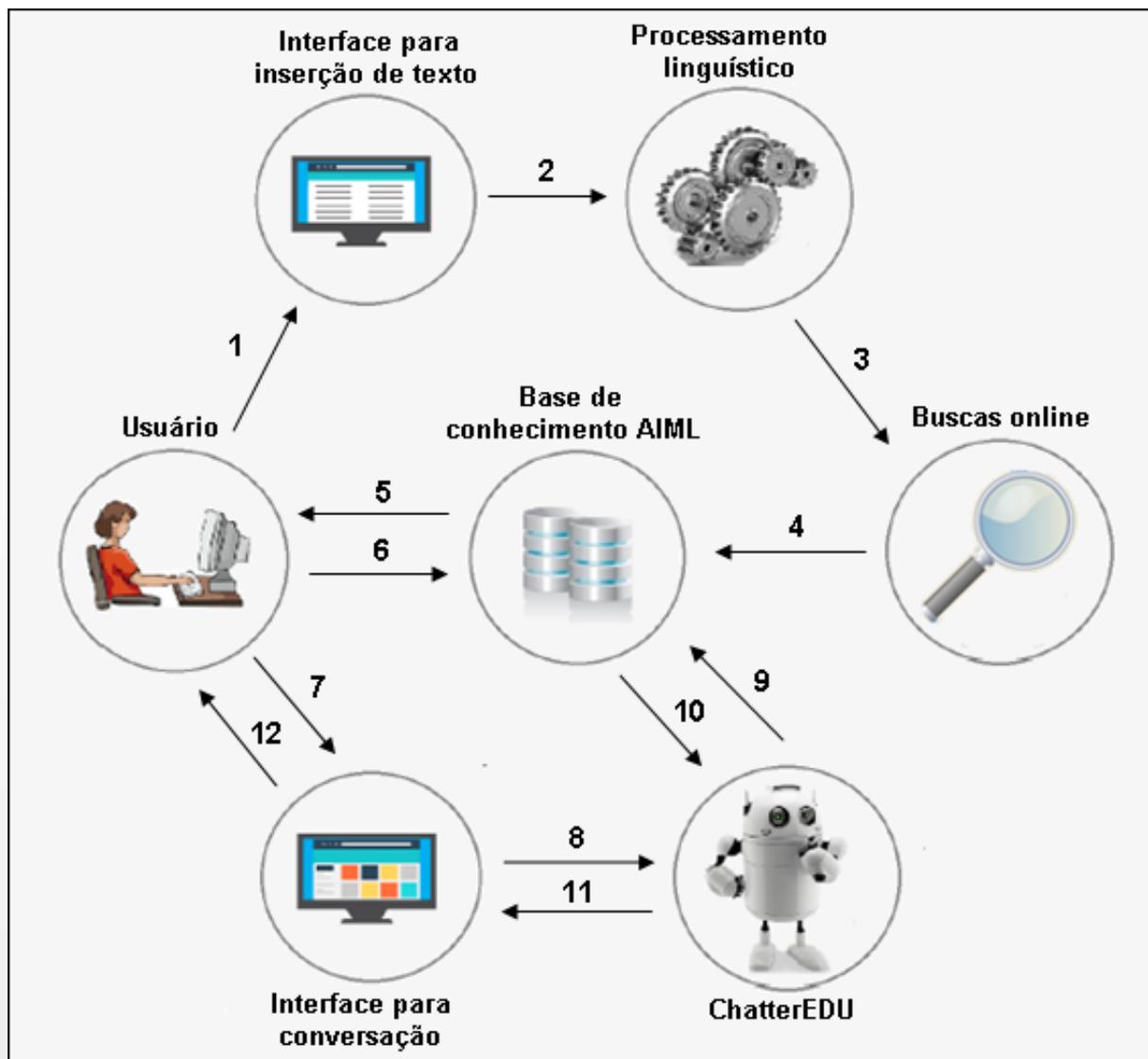


Diagrama de casos de uso

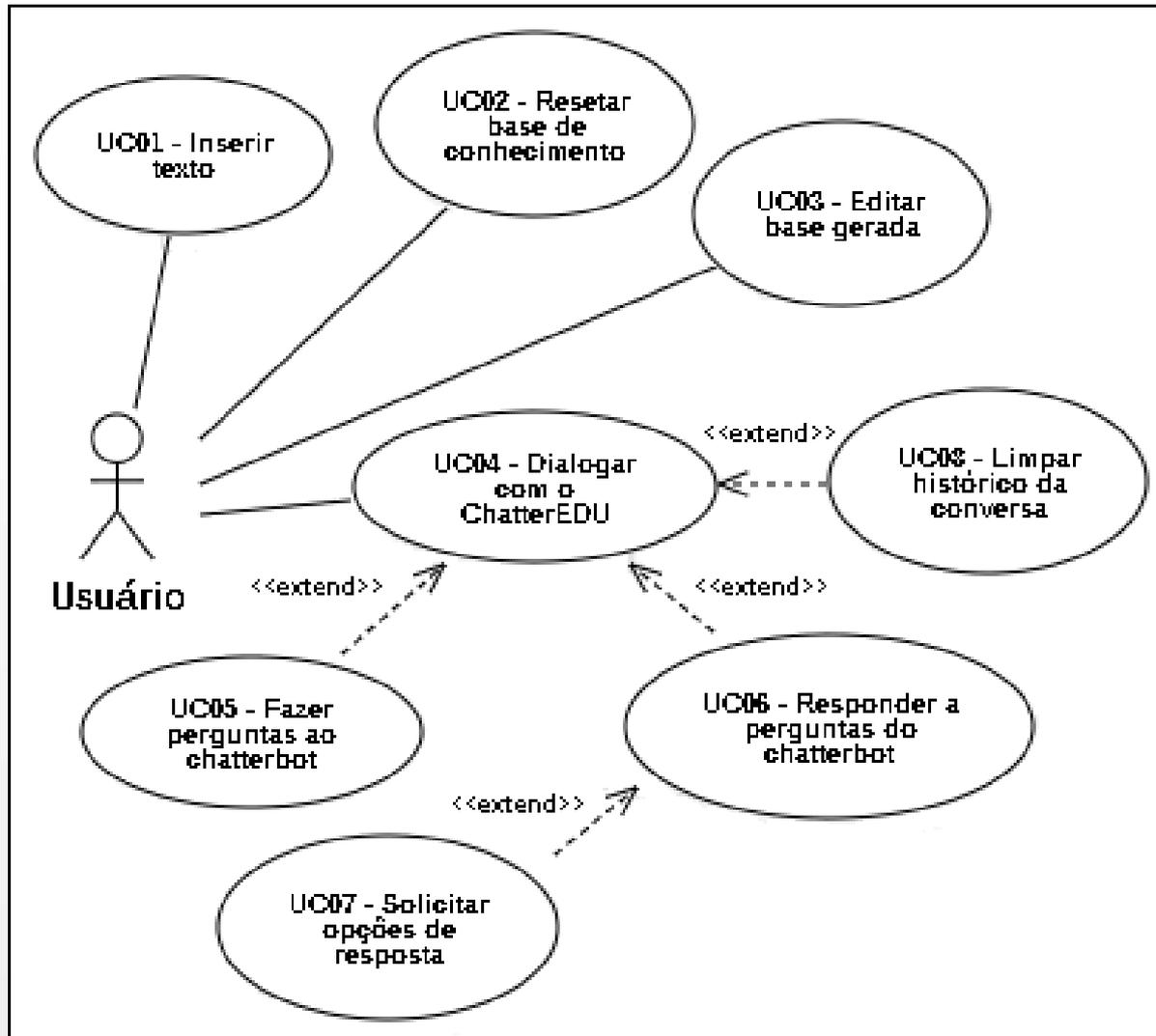


Diagrama de casos de uso

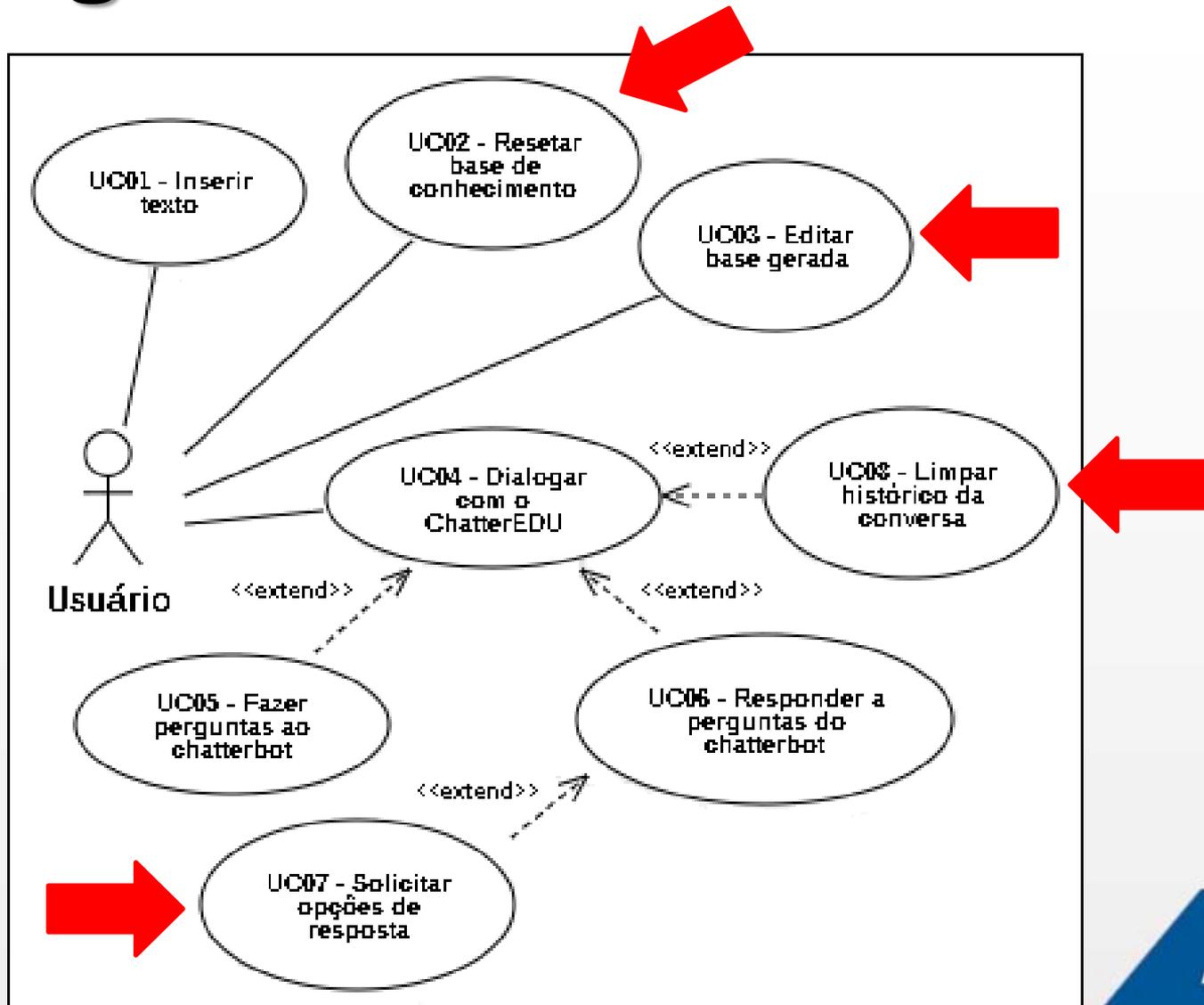


Diagrama de classes

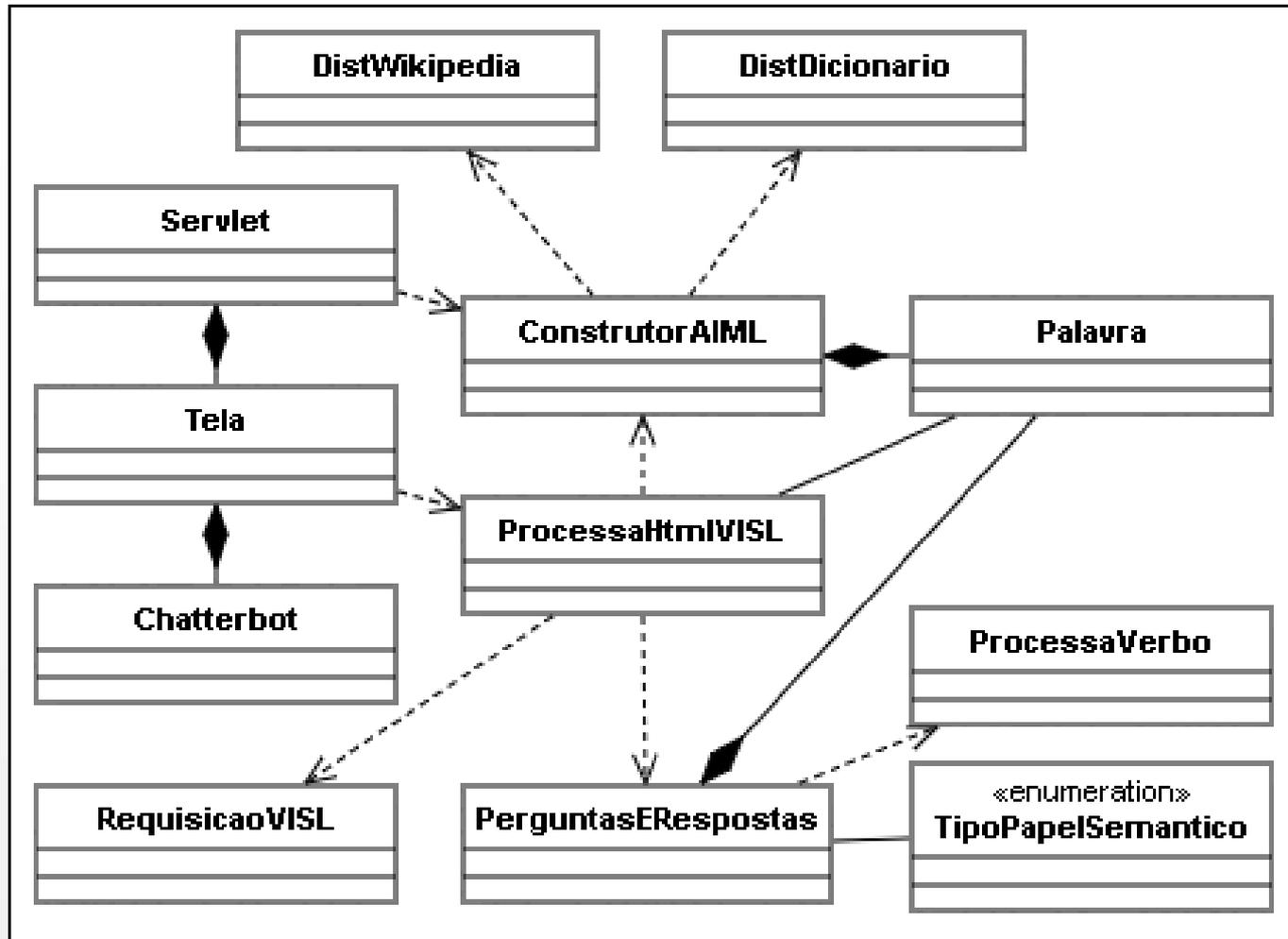


Diagrama de classes

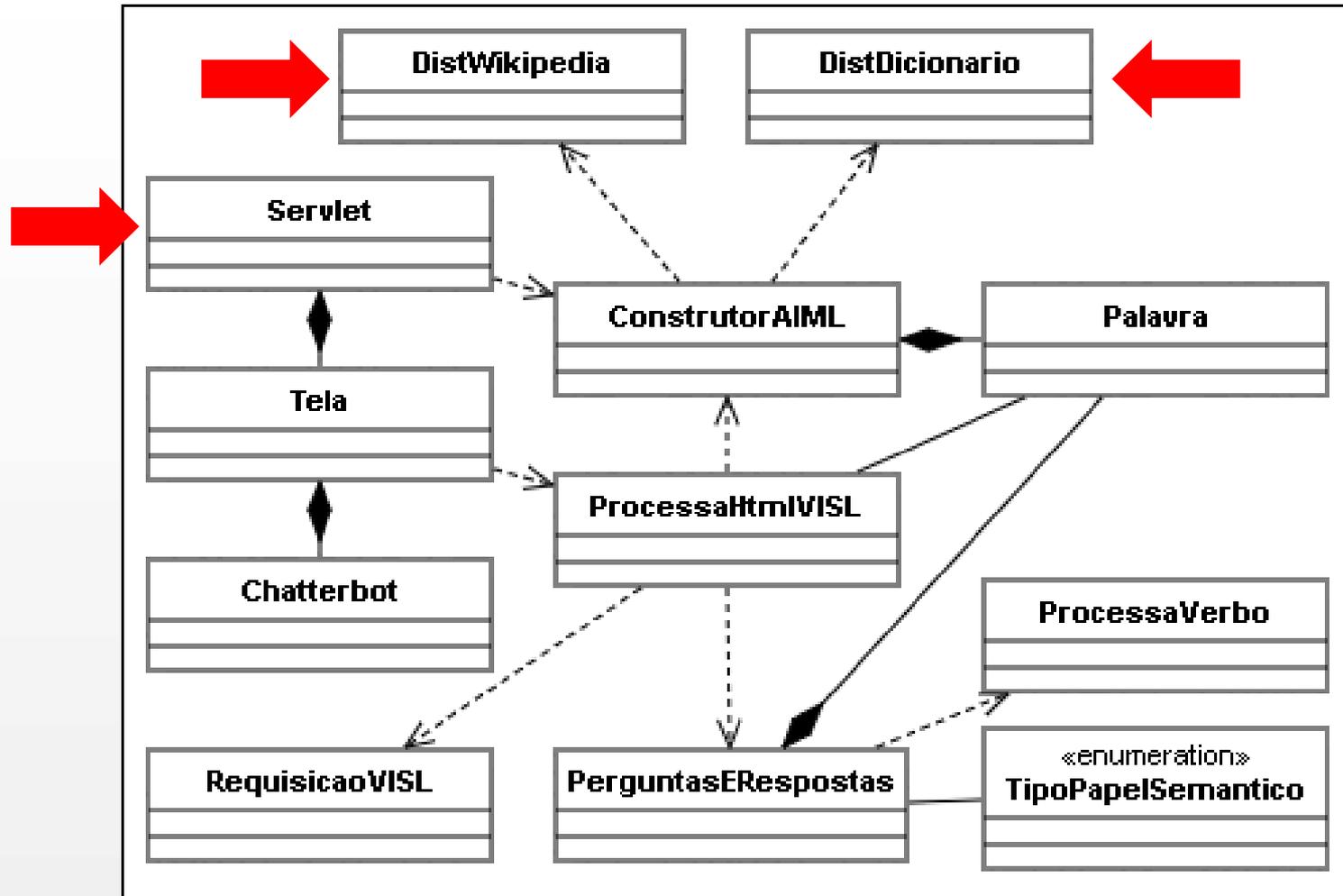


Diagrama de classes

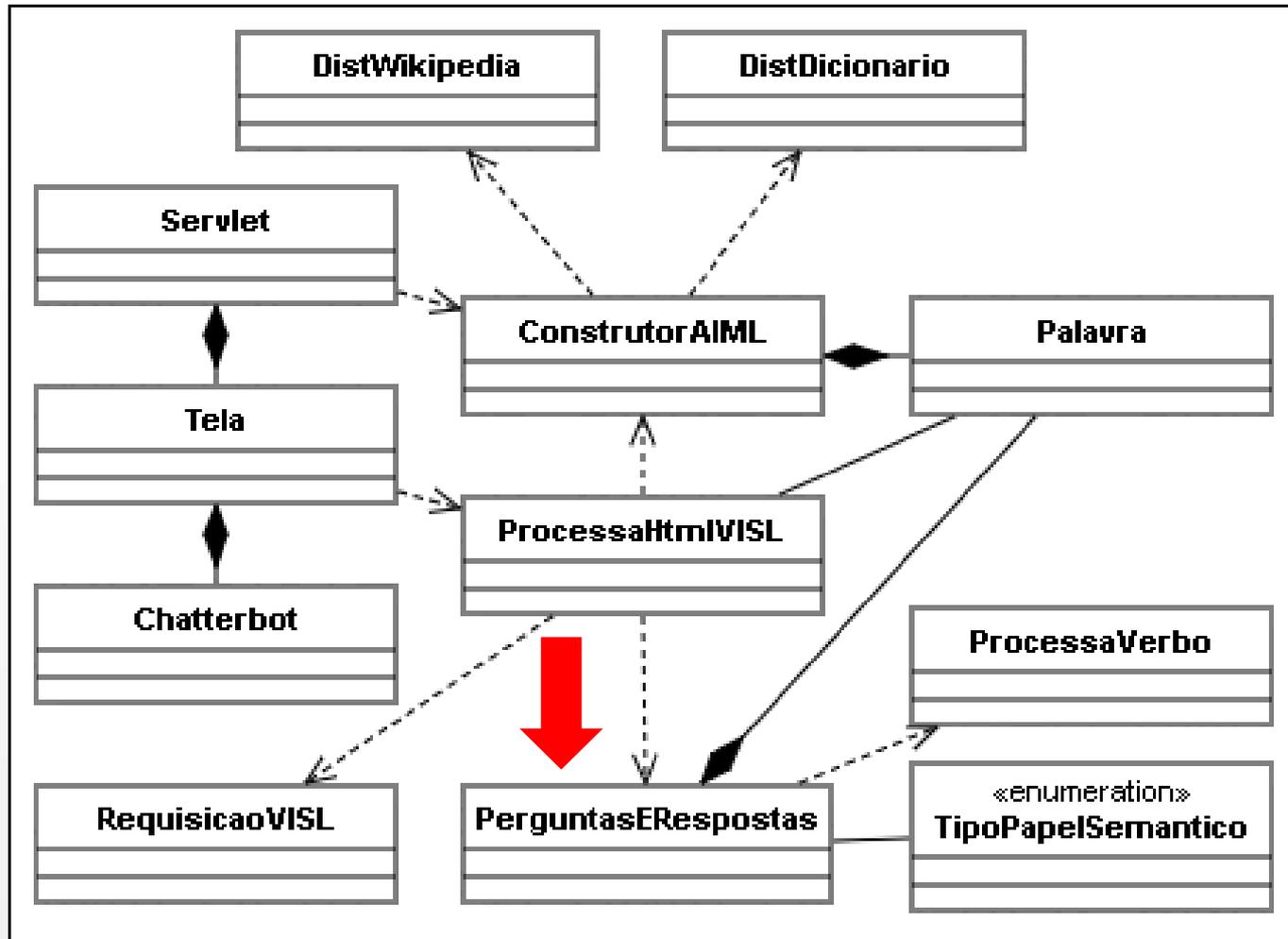


Diagrama de classes

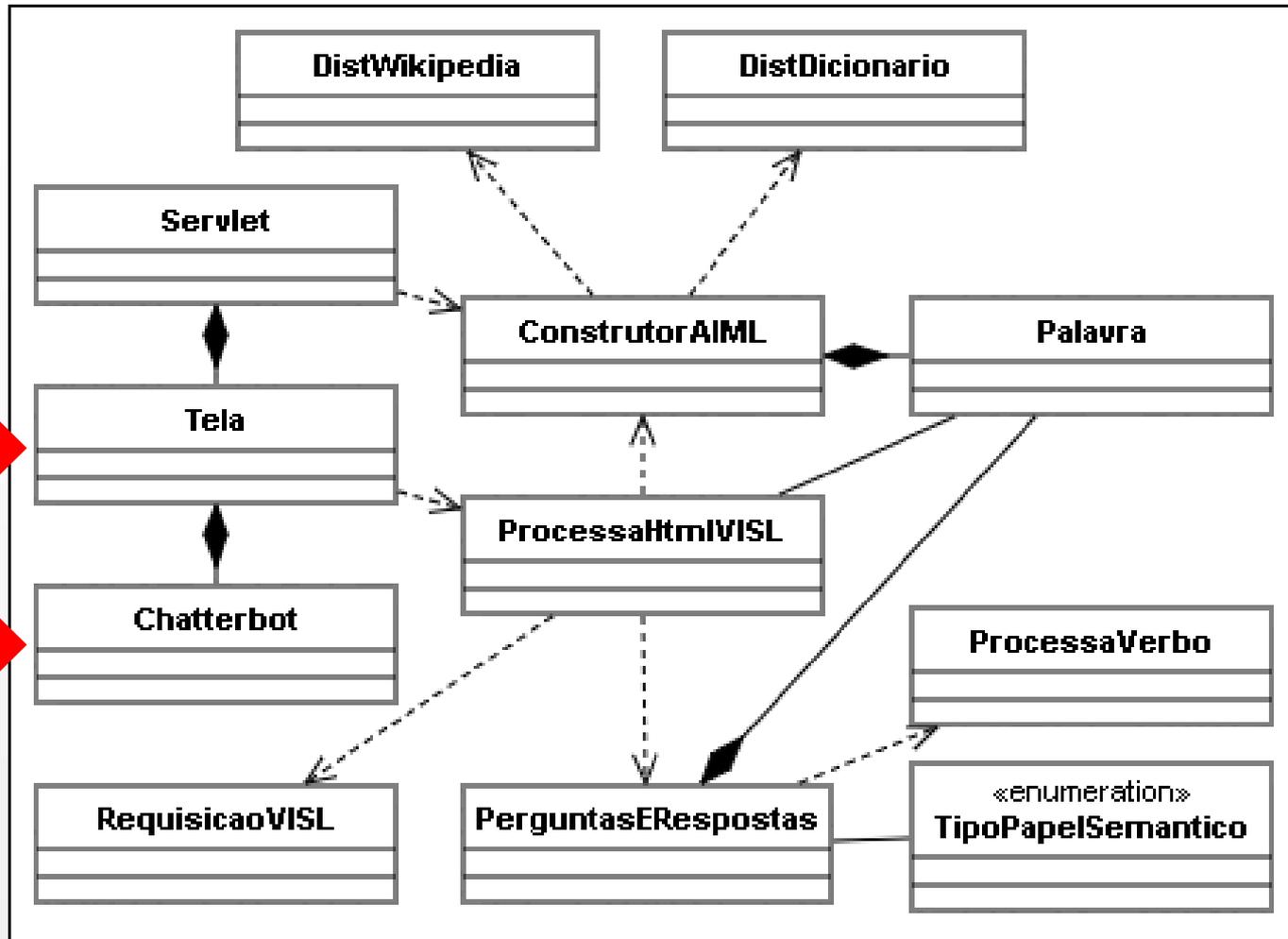


Diagrama de classes

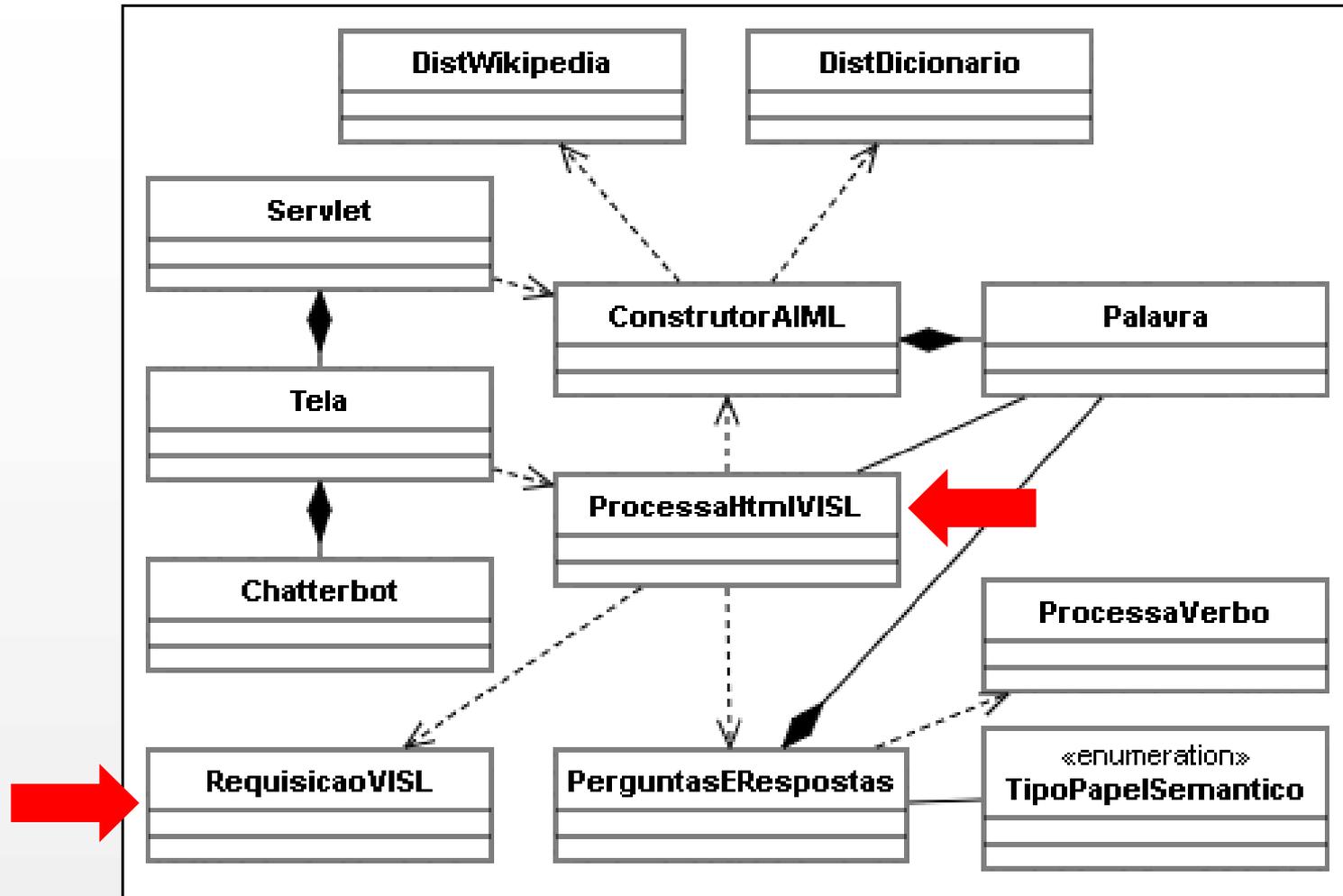
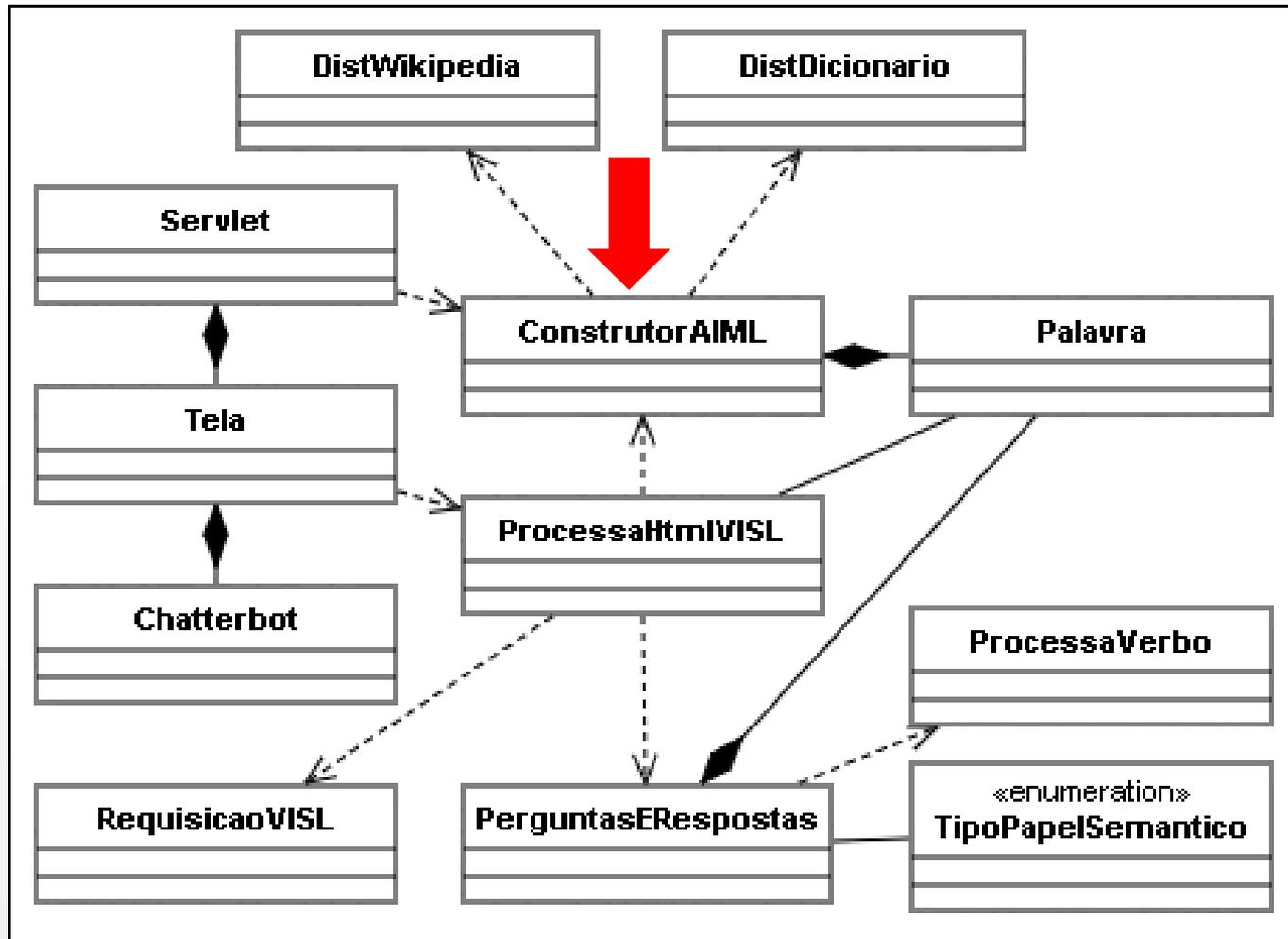


Diagrama de classes



Implementação

- Geração de distratores:
 - Respostas com numeral;
 - Respostas sem numeral.

Respostas com numeral

- Identificação da classe sintática NUM:

```
O [o] <*> DET M S @>N #1->2
Brasil [Brasil] <*> PROP M S @SUBJ> §PAT #2->3
foi [ser] <fmc> V PS 3S IND VFIN @FS-STA #3->0
descoberto [descobrir] V PCP PAS M S @ICL-AUX< §PRED #4->3
em [em] PRP @<ADVL #5->4
1500 [1500] <card> NUM M/F P @P< §LOC-TMP #6->5
.[.] PU @PU #7->0 </β> </β>
```

Respostas com numeral

- Pergunta: Quando o Brasil foi descoberto?
- Resposta: Em 1500.

- Ocorre a substituição do termo “1500” por outros números.

Respostas com numeral

- Valores entre 1.000 e 2.050:
 - Geração de oito resultados randômicos entre 10 unidades abaixo e 10 unidades acima.
- Valores em demais intervalos:
 - Geração de um décimo, um quarto, metade, três quartos, cinco quartos, dobro, quádruplo e décuplo.

Respostas com numeral

- Quando o Brasil foi descoberto?
- Em 1500. (correta)
- Em 1498.
- Em 1492.
- Em 1509.
- Em 1507.
- Em 1495.
- Em 1503.
- Em 1497.
- Em 1496.

Respostas com numeral

- Sentença: o barril de petróleo vale 80 dólares.
 - O barril de petróleo vale quanto?
 - 80 dólares. (correta)
 - 8 dólares.
 - 20 dólares.
 - 40 dólares.
 - 60 dólares.
 - 100 dólares.
 - 160 dólares.
 - 320 dólares.
 - 800 dólares.

Respostas sem numerais

- Acesso à Wikipédia e ao Dicionário Criativo:
 - <https://dicionariocriativo.com.br/> + resposta;
 - <https://pt.wikipedia.org/w/index.php?search=> + resposta.

Respostas sem numerais



planalto



Resumo

Brainstorm

Significado

Sinônimos

Relacionadas

Expressões

Citações

🔗 PALAVRAS RELACIONADAS

planalto

outeiro

montanha

serrania

serra

montículo

monte

cordilheira

colina

promontório

morro

meseta

altura

alteza

elevação

altitude

eminência

alto

sublimidade

superioridade

protuberância

grandeza

estatura

pujança

pico

cume

pinCARO

pináculo

topo

cimo

[+relacionadas a planalto](#)

Respostas sem numerais

```
1 <[...] title="planalto" class="c_primary_hover c_border_hover analogico">
2     planalto
3 <[...]>
4 <[...] title="outeiro" class="c_primary_hover c_border_hover analogico">
5     outeiro
6 <[...]>
7 <[...] title="montanha" class="c_primary_hover c_border_hover analogico">
8     montanha
9 <[...]>
10 <[...] title="serrania" class="c_primary_hover c_border_hover analogico">
11     serrania
12 <[...]>
13 <[...] title="serra" class="c_primary_hover c_border_hover analogico">
14     serra
15 <[...]>
```

Respostas sem numerais

```
1 <[...] title="planalto" class="c_primary_hover c_border_hover analogico">
2     planalto
3 [...]>
4 <[...] title="outeiro" class="c_primary_hover c_border_hover analogico">
5     outeiro
6 [...]>
7 <[...] title="montanha" class="c_primary_hover c_border_hover analogico">
8     montanha
9 [...]>
10 <[...] title="serrania" class="c_primary_hover c_border_hover analogico">
11     serrania
12 [...]>
13 <[...] title="serra" class="c_primary_hover c_border_hover analogico">
14     serra
15 [...]>
```

Respostas sem numerais

```
1 <[...] title="planalto" class="c_primary_hover c_border_hover analogico">
2   planalto
3 [...]>
4 <[...] title="outeiro" class="c_primary_hover c_border_hover analogico">
5   outeiro
6 [...]>
7 <[...] title="montanha" class="c_primary_hover c_border_hover analogico">
8   montanha
9 [...]>
10 <[...] title="serrania" class="c_primary_hover c_border_hover analogico">
11   serrania
12 [...]>
13 <[...] title="serra" class="c_primary_hover c_border_hover analogico">
14   serra
15 [...]>
```

Respostas sem numerais

```
1 <[...] title="planalto" class="c_primary_hover c_border_hover analogico">  
2   planalto  
3 <[...]>  
4 <[...] title="outeiro" class="c_primary_hover c_border_hover analogico">  
5   outeiro  
6 <[...]>  
7 <[...] title="montanha" class="c_primary_hover c_border_hover analogico">  
8   montanha  
9 <[...]>  
10 <[...] title="serrania" class="c_primary_hover c_border_hover analogico">  
11   serrania  
12 <[...]>  
13 <[...] title="serra" class="c_primary_hover c_border_hover analogico">  
14   serra  
15 <[...]>
```

Respostas sem numerais

Artigo Discussão

Blumenau

Origem: Wikipédia, a enciclopédia livre.

Nota: Par

Blumenau é um
Mesorregião do V
do estado, a 8ª d
seus principais p

Blumenau

 **Portal do Brasil** 

Controle de autoridade  : Q155557 • VIAF: 153611830 • EBID: ID • GEC: 0010567 • GND: 4007236

Categorias: [Blumenau](#) | [Fundações no Brasil em 1850](#)

```
<a href="/wiki/Categoria:Blumenau" title="Categoria:Blumenau">Blumenau</a>
```

`/wiki/Categoria: + resposta`

Respostas sem numerais

Categoria:Blumenau

Origem: Wikipédia, a enciclopédia livre.

Subcategorias

Esta categoria contém as seguintes 14 subcategorias (de um total de 14).

!

- ▶ [!Predefinições sobre Blumenau](#) (4 P)

A

- ▶ [Atrações turísticas de Blumenau](#) (1 C, 6 P)

C

- ▶ [Construções de Blumenau](#) (1 C, 7 P)
- ▶ [Cultura de Blumenau](#) (2 C, 6 P)

E

- ▶ [Economia de Blumenau](#) (2 C, 1 P)
- ▶ [Educação em Blumenau](#) (1 C, 4 P)
- ▶ [Esporte de Blumenau](#) (2 C, 12 P)

G

- ▶ [Geografia de Blumenau](#) (2 C, 10 P)

H

- ▶ [História de Blumenau](#) (8 P)

N

- ▶ [Natu](#)

[Categoria: Municípios de Santa Catarina](#)

Respostas sem numerais

Categoria:Municípios de Santa Catarina

Origem: Wikipédia, a enciclopédia livre.

Subcategorias

Esta categoria contém as seguintes 200 subcategorias (de um total de 216).

([página anterior](#)) ([página seguinte](#))

▶ [Listas de municípios de Santa Catarina](#) (7 P)

I

▶ [IPredefinições sobre municípios de Santa Catarina](#) (1 C, 2 P)

A

▶ [Abdon Batista](#) (3 C, 1 P)

▶ [Abelardo Luz](#) (5 C, 2 P)

▶ [Agrolândia](#) (5 C, 1 P)

▶ [Agronômica](#) (3 C, 1 P)

▶ [Água Doce](#) (2 C, 1 P)

▶ [Águas Frias \(Santa Catarina\)](#) (1 C, 1 P)

F

▶ [Faxinal dos Guedes](#) (2 C, 1 P)

▶ [Florianópolis](#) (12 C, 14 P)

▶ [Forquilha](#) (4 C, 1 P)

▶ [Fraiburgo](#) (4 C, 1 P)

G

▶ [Galvão \(Santa Catarina\)](#) (2 C, 1 P)

▶ [Garopaba](#) (5 C, 1 P)

▶ [Garuva](#) (3 C, 1 P)

▶ [Gaspar](#) (5 C, 1 P)

▶ [Governador Celso Ramos](#) (4 C, 1 P)

▶ [Grão Pará](#) (5 C, 1 P)

▶ [Gravatal](#) (5 C, 1 P)

Respostas sem numerais

Artigo	Discussão
<h2>Campo Erê</h2> <hr/>	
<p>Origem: Wikipédia, a enciclopédia livre.</p>	
<p>Campo Erê é um município da Mesorregião do Oeste estimada em 2004 era de 10 437 habitantes.</p>	
<p>O município é banhado pelos rios: Sa</p>	
<p>Categoria: Municípios de Santa Catarina</p>	

Demonstração da Implementação

Resultados e Discussões

- Objetivos atingidos:
 - Geração de questões de múltipla escolha com busca por distratores em páginas virtuais;
 - Migração do *chatterbot* para a plataforma web;
 - Ajustes nos tratamentos de 6 dos 10 papéis semânticos.

Resultados e Discussões

- Testes realizados:
 - 42 sentenças, divididas em três assuntos:
 - História/Geografia;
 - Informática;
 - Esportes.
 - Foram gerados 68 conjuntos perguntas/respostas/distratores.
 - 63% avaliados como bons;
 - 34% avaliados como satisfatórios;
 - 3% avaliados como ruins.

Resultados e Discussões

- Pergunta avaliada como boa:
- Quem possui 27 unidades federativas?
 - O Chile
 - O Suriname
 - O Venezuela
 - O Equador
 - O Brasil

Resultados e Discussões

- Pergunta avaliada como satisfatória:
- Qual é Estado-membro da União Europeia desde 2004?
 - A Noruega
 - A Liechtenstein
 - A Áustria
 - A Suíça
 - A Eslováquia

Resultados e Discussões

- Pergunta avaliada como ruim:
- Onde a Rússia faz?
 - A maior de suas fronteiras com a Camboja
 - A maior de suas fronteiras com a Maldivas
 - A maior de suas fronteiras com a Egito
 - A maior de suas fronteiras com a Timor-Leste
 - A maior de suas fronteiras com a Mongólia

Resultados e Discussões

- O maior número de frases tende a aumentar o tempo de processamento:
 - Textos de 160 palavras:
 - 16 frases: 107s;
 - 24 frases: 166s.

Resultados e Discussões

- Frases longas prejudicam o tratamento:
 - Textos de 160 palavras:
 - 16 frases: 12% avaliadas como ruins;
 - 24 frases: 6% avaliadas como ruins.

Resultados e Discussões

- Problemas de concordância quando apenas parte da resposta é extraída.
- A capital do Brasil:
 - A capital do Uruguai;
 - A capital do Colômbia.

Resultados e Discussões

- Foi desenvolvido um tratamento para as inconsistências:
 - Quadruplicou o tempo de processamento;
 - Inconsistências persistiam.

Resultados e Discussões

- Uma máquina universal:
 - **Umas** neurociências Forenses universal;
 - **Umo** veículo universal.

Resultados e Discussões

- Melhoria do tratamento LOC-TMP:
- **Sentença: O Brasil foi descoberto no mês de abril.**

Resultados e Discussões

- Processamento do ChatterEDU 1.0:
 - Pergunta: Quando o Brasil foi descoberto abril
 - Resposta: No mês de

Resultados e Discussões

- Processamento do ChatterEDU 1.0:
 - Pergunta: Quando o Brasil foi descoberto abril
 - Resposta: No mês de
- Processamento do ChatterEDU 2.0:
 - Pergunta: Quando o Brasil foi descoberto
 - Resposta: No mês de abril

Resultados e Discussões

- Melhoria do tratamento LOC:
- **Sentença: A maioria das favelas não dispõe de infraestrutura básica.**

Resultados e Discussões

- Processamento do ChatterEDU 1.0:
 - Pergunta: Onde dispõe de infraestrutura básica
 - Resposta: A maioria de as favelas não

Resultados e Discussões

- Processamento do ChatterEDU 1.0:
 - Pergunta: Onde dispõe de infraestrutura básica
 - Resposta: A maioria de as favelas não
- Processamento do ChatterEDU 2.0:
 - Pergunta: Onde não se dispõe de infraestrutura básica
 - Resposta: Na maioria das favelas

Resultados e Discussões

características	ChatterEDU 2.0	Brito (2017)	Souza e Moraes (2015)	Curto (2010)
correlatos				
plataforma	web	web	web	não descrito
gera perguntas ou respostas automaticamente a partir de um texto de entrada ou página virtual	perguntas e resposta	respostas	respostas	perguntas e respostas
fornece respostas a perguntas feitas pelo usuário	sim	sim	sim	não
permite que sejam feitas perguntas ao usuário	sim	não	não	sim
gera distratores para as perguntas	sim, a partir de páginas virtuais	não	não	sim, a partir do texto de entrada
é voltado à educação	sim	não	não	sim

Conclusões

- Objetivo inicial de limitar o assunto a Geografia de ensino básico:
 - O sistema mostrou-se adequado para qualquer assunto;

Conclusões

- Benefícios da geração de distratores via páginas virtuais;
- Contribuições para área educacional e de entretenimento.

Extensões

- Hospedar a aplicação em um domínio;
- Desenvolver tratamentos em outros idiomas;
- Aperfeiçoar os tratamentos de papéis semânticos já existentes:
 - reduzir as inconsistências remanescentes;

Extensões

- persistir a base AIML através de um Sistema Gerenciador de Banco de Dados (SGBD);
- reduzir o tempo de busca por distratores por meio de novas técnicas;

Extensões

- incluir o processamento para frases com outras estruturas gramaticais:
 - não apenas na ordem sujeito, verbo e objeto.

Obrigado!