

# Plano de Doutoramento

---

Construção de uma Base de  
Conhecimento Lexical para o  
Português

Nuno Seco

# Estrutura da Apresentação

---

- ❑ Objectivos
  - ❑ Motivação
  - ❑ Expectativas
  - ❑ Plano
  - ❑ Critérios de Sucesso
-

# Objectivo

---

- Criar uma Base de Conhecimento Lexical para o Português.
    - Formalização de uma metodologia
    - Uma implementação da metodologia
-

# Objectivo

---

- Tornar explícita a informação sobre conceitos (referenciados por termos).
    - O que é **árvore**?
      - *“planta alta de caule (tronco) e ramos lenhosos”*
    - *Para que servem? São seres vivos? Se sim, como nascem? Quantos tipos existem?*
-

# Definição de Base de Conhecimento Lexical

---

“É um repositório de informação sobre conceitos que pode ser utilizado em áreas como a Linguística, IA e Ciências da Informação (...) contém informação sobre conceitos lexicais e as suas relações.”

Robert Amsler (1984)

---

# Definição de Base de Conhecimento Lexical

---

“É uma estrutura que fornece uma **organização semântica** aos termos do nosso léxico, tornando **explícita** as relações semânticas entre estes.”

---

# Motivação

---

- ❑ O nosso sistema conceptual está organizado como uma rede de conceitos. (Miller, Quillan)
  - ❑ O sistema conceptual governa a forma como: pensamos, agimos e comunicamos. (Lackoff)
  - ❑ Conceitos esses que são referenciados por termos.
-

# Motivação

---

Para as máquinas:

- Recolha de Informação
- *Desambiguação Semântica e Sintáctica*
- *Semelhança semântica entre Palavras*
- *Pergunta e Resposta*

□ Para os humanos:

- Pode servir como um auxiliar de escrita, como um Dicionário ou Tesouro.
-



# Abordagens Estatísticas

---

- Algoritmos estatísticos, como frequências de co-ocorrências, são semanticamente cegos.
    - carro
    - pneu
    - volante
    - motor
    - veículo
    - condutor
-

# Abordagens Manuais

---

- Conhecimento é compilado manualmente.
    - WordNet, Cyc, HowNet
  
  - Problemas:
    - Muito moroso
    - Muito caro
-

# Abordagens Baseadas em Extracção de Informação

---

- Procurar padrões que indicam a existência de uma relação semântica no texto:

parte\_de(pneu, carro)

tipo\_de(carro, veículo)

parte\_de(motor, veículo)

usa(condutor, carro)

---

# Abordagem Proposta

---

- Baseada em Extracção de Informação.
    - Análise das definições em dicionários.  
(Inspirada na abordagem do MindNet)
-

# Abordagem Proposta

---

- Informação sintáctica
    - Substantivos
    - Verbos
    - Advérbios
    - Adjectivos
  
  - Informação predicativa
    - comida é o **objecto\_típico** de verbo comer
-

# Abordagem Proposta

---

- Informação sobre os diferentes sentidos de um termo
    - Um representação para lidar com *homonímia*:
      - parte\_de(banco.1, jardim)
      - instância\_de(BES, banco.2)
-

# Exemplo

---

CARRO.1

Categoria: Substantivo

Definição:

Hiperónimos: veículo.1

Hipónimos:

Merónimos: pneu, motor, volante

Objecto\_Típico\_de: conduzir.2

(...)

---

# Expectativas

---

- Existência de uma Ontologia Lexical para o Português.
  - Um sistema informático (semi-)automático de extracção de conhecimento.
    - Capaz de Evoluir
      - Adição de novo conhecimento
      - Refinamento do existente
-



# Exemplo de utilização

---

- ❑ Sistema de Pergunta e Resposta:

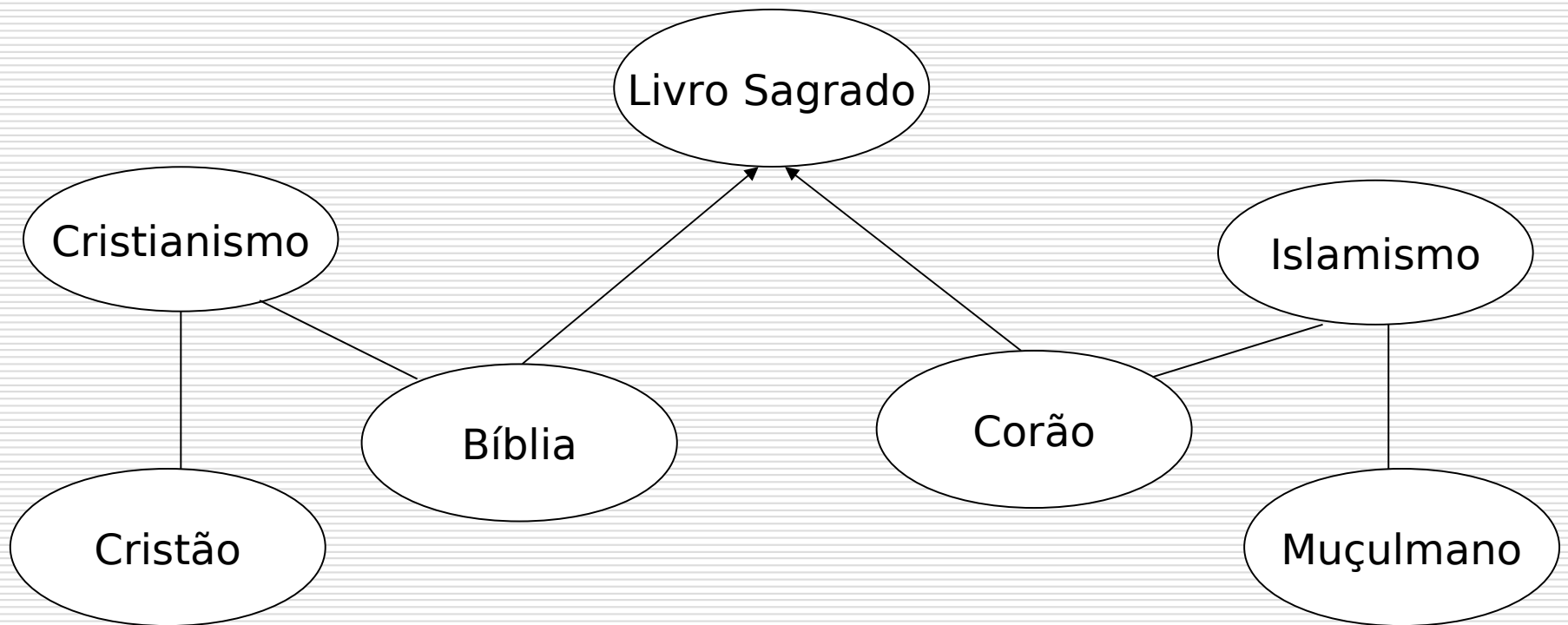
```
oraculo:\> Existe alguma bíblia muçulmana?
```

```
oraculo:\> Corão, Alcorão
```

---

# Exemplo de Utilização

---



# Planificação

---

- Revisão Bibliográfica.
    - Identificação das relações semânticas mais importantes.
  - Procura de recursos para iniciar o processo de extracção. (Dicionários)
  - Definição do modelo de dados para armazenamento.
-

# Planificação

---

- Protótipo do sistema de extracção.
    - Formalização do Processo de Extracção
    - Base de Conhecimento Lexical.
  - Análise do recurso gerado.
    - Aplicação em tarefas específicas.
    - Medição do Impacto
  - Melhoramentos.
-

# Critérios de Sucesso

---

- Utilização do recurso por terceiros
  
  - Contribuição do recurso em tarefas relacionadas com PLN:
    - Recolha de Informação
    - Desambiguação
    - Semelhança Semântica
    - Sistemas de Pergunta e Resposta
-

# Plano de Doutoramento

---

Construção de uma Ontologia  
Lexical para o Português

Nuno Seco