**Arquitetura para um sistema de perguntas e respostas usando ontologias**

Brian Alves Andreossi 11060215 Gustavo Zanfelice Dib 11023915 Marcelo Schirbel Gomes 11022014 Murilo Bolzan Dionisio 11107414

**Artigo utilizado:** Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System

**Athira P. M., Sreeja M. and P. C. Reghuraj**

*Department of Computer Science and Engineering, Government Engineering College, Sreekrishnapuram, Palakkad Kerala, India*

**Novembro de 2013**

● **19** Citações

● **11** Referências

**● 10** Páginas

**Ideia Geral**

**Uma arquitetura para um sistema de Perguntas e respostas que analisasse a ontologia das perguntas**

Arquitetura Comum:

Arquitetura Proposta:

**Arquitetura Proposta**

**Adaptação do Problema**

**Sistema de resposta automática baseado no Stack Overflow**

Base de dados:

Bases:

Question.csv

Base de perguntas com Id, Score, Texto e corpo. Tags.csvBase de tags com Id da pergunta e Tags Answers.csv

Base de Respostas com Id da resposta, Id da pergunta, Score e corpo da resposta

**Question Processing**

**Modelo de Classificação**

Como temos a base de Tags, utilizamos métodos de Aprendizado supervisionado para classificar uma nova pergunta com as tags do StackOverflow.

O método de melhor desempenho foi o Naive Bayes multinomial no paradigma One vs. All, para realizarmos a classificação com diversas Classes.

**Exee resposta**

**Document Processing**

**Distância entre as perguntas**

Usamos a distância de Levenshtein para obter uma aproximação entre todas as palavras das frases e depois a distância do cosseno para comparar as frases inteiras, por fim, somamos os dois valores.

**Answer Extraction**

**Extraindo Perguntas**

A base de busca é filtrada pelas tags preditas. O retorno dessa função é um dicionário com as melhores perguntas, 5 melhores. Caso o usuário não insira nenhuma pergunta das selecionadas, retornamos as 5 próximas seguintes.

**Perguntas do Professor**

**Por que é necessário um sistema de perguntas e respostas?**

**O primeiro ponto é termos a capacidade de obter a resposta mesmo offline.**

**O segundo ponto é que obtemos a melhor resposta da base de dados, sem depender encontrar uma thread certeira.**

**Quais técnicas de PLN utilizamos?**

**Naive Bayes Multinomial(one versus all), Distância de Levenshtein, Distância de Cosseno,**

**Quais os sistemas de perguntas e respostas conhecidos atualmente? Qual tecnologia eles usam?**

**Quora, Google Answers e Yahoo Answers!**

**Algumas das tecnologias estão descritas no livro Introduction to Information Retrieval, disponível no seguinte link:** https://nlp.stanford.edu/IR-book/

**Estão inclusas: Naive Bayes, Flat Clustering, Regex e Probabilistic Information Retrieval**

**Como é feita a validação da proposta?**

**Se a dúvida do usuário for sanada.**

**Para fazermos uma validação mais a fundo, podemos fazer uma pesquisa em outros sites de perguntas e respostas.**