

UTILIZAÇÃO DE PROCESSAMENTO DE LINGUAGEM NATURAL E ONTOLOGIAS NA ANÁLISE QUALITATIVA DE FRASES CURTAS

Felipe Ribeiro Pereira – UNISINOS – feliperibeirópereira@gmail.com

Sandro Jose Rigo – UNISINOS – rigo@unisinós.br

Resumo: O volume de dados digitais disponíveis é cada vez maior em todos os segmentos profissionais e essa realidade não é diferente no campo educacional, principalmente quando falamos sobre Educação a Distância e ambientes virtuais de aprendizagem. Um dos desafios neste cenário é o tratamento de dados em linguagem natural, de forma com que existam ganhos de qualidade no processo de ensino-aprendizagem. O objetivo do presente estudo é apresentar e avaliar um modelo para a análise qualitativa de respostas em linguagem natural, provenientes de interações em um ambiente virtual de aprendizagem. Para isso, também são utilizadas técnicas de processamento de linguagem natural e representação de conhecimento em ontologias. A avaliação preliminar realizada com a abordagem proposta, através de um protótipo desenvolvido, indica possibilidades de bons resultados e a possibilidade de automatização do processo, o que pode facilitar o trabalho de educadores que precisam se envolver com um alto volume de informações para avaliar um grande número de estudantes, bem como pode ser útil para estudantes em processos de estudo individual.

Palavras-chave: *Ontologia, Processamento de Linguagem Natural. Ambiente Virtual de Aprendizagem.*

Abstract: The volume of available digital data is growing in all professional segments and that reality is not different in the educational field, especially when we talk about Distance Education and virtual learning environments. One of the challenges in this scenario is the processing of natural language, in order that there are quality gains in the process of teaching and learning. The purpose of this study is to present and evaluate a model for the qualitative analysis of answers in natural language, originating from interactions in a virtual learning environment. For this, techniques of natural language processing and representation of knowledge in ontologies are also used. The evaluation of the proposed approach, with a prototype, indicates possibilities of good results and automation of the process, which can facilitate the work of educators who need to be involved with a high volume of information to evaluate a large number of students, and can be useful for students in processes of individual study.

Keywords: *Ontology. Natural Language Processing. Virtual Learning Environment*

1 INTRODUÇÃO

Estamos vivendo a era da informação e as organizações estão sendo guiadas por ela. A quantidade de dados disponíveis é muito grande em todos os segmentos organizacionais e cada vez mais se torna indispensável agrupar e transformar estes dados em informações relevantes para o apoio à tomada de decisões. No campo educacional podemos observar exatamente a mesma situação, principalmente quando falamos sobre os Ambientes Virtuais de Aprendizagem (AVAs). Este novo formato de interação social gera um volume grande de informações de texto em linguagem natural, concentradas em chats, fóruns e outras ferramentas de interação.

Devido ao aumento do volume de dados citados acima, entende-se que é uma tarefa cada vez mais difícil o papel de analisar todas estas informações para realizar avaliações e interagir com os alunos quando necessário. Cabe destacar neste contexto,

portanto, a possibilidade de utilização de recursos digitais para atividades de apoio ao professor na correção de exercícios dos alunos. Diante deste cenário, identificou-se como uma oportunidade pesquisar o emprego de tecnologias de Processamento de Linguagem Natural (PLN) (Liddy, 2003), juntamente com o uso de representação de conhecimento a partir de ontologias (Grubber, 2003) para gerar informações sumarizadas sobre a correção de respostas em linguagem natural, a partir de um conjunto de dados extraídos de um AVA, com o foco na avaliação da qualidade dos textos de respostas.

Grande parte dos estudos realizados atualmente nesta área é baseada em construções matemáticas para modelar os relacionamentos existentes nos textos tratados. As técnicas consistem, principalmente, na análise do número de incidências e da proximidade das palavras em uma mensagem. Neste contexto, podemos citar os trabalhos de Gazolla (2011) e Azevedo (2011), nos quais são utilizadas estas técnicas para identificação de padrões nos comentários inseridos por alunos em fóruns de aprendizado. O trabalho de Gazolla estuda a polaridade de opiniões, e para isso, analisa a polaridade dos adjetivos presentes nos textos. Já o estudo de Azevedo é focado na análise de relevância dos textos em comparação a um determinado tema, classificando-os com um determinado fator previamente cadastrado.

Mesmo sendo técnicas interessantes e de uso apropriado para diversas áreas do conhecimento, quando inseridas no contexto educacional elas podem gerar resultados indesejados. Quando objetivamos avaliar a qualidade de respostas a respeito de um determinado conceito, precisamos não só avaliar a presença de palavras-chave, mas também as avaliações de outros aspectos importantes se tornam necessárias, tais como o estilo da escrita, o uso de sinônimos, o emprego de termos polissêmicos, o contexto das frases, entre outros.

Este trabalho tem o objetivo de propor e avaliar alternativas ao problema detalhado acima, pois melhorias nestas ferramentas podem ser de grande utilidade na área da educação, entre outras. Para isso, será avaliado o uso integrado de PLN e de ontologias no apoio às avaliações de respostas curtas, em frases na linguagem natural. O objetivo central da solução proposta é o apoio na avaliação de frases curtas contendo respostas de alunos, provenientes de interações em um ambiente virtual de aprendizagem. A avaliação das frases é realizada com apoio de recursos de PLN e regras linguísticas que permitem a identificação dos elementos importantes das respostas, que são então consultadas em uma ontologia de domínio para identificar sua correção.

O texto deste artigo está estruturado de forma a apresentar inicialmente os trabalhos e artigos relacionados, seguidos da apresentação da proposta de arquitetura, resultados, e, por fim, as conclusões e considerações finais.

2 TRABALHOS RELACIONADOS

Este capítulo abordará alguns estudos recentes relacionados ao tema estudado. Sendo assim, procurou-se identificar exemplos de soluções já desenvolvidas para ambientes virtuais de aprendizagem e estudos relacionados à utilização de ontologias e PLN em outras áreas do conhecimento.

2.1 MINERAFÓRUM: Um recurso de apoio para análise qualitativa em fóruns de discussão

A tese de doutorado de Azevedo (2011) é dedicada ao desenvolvimento de recursos tecnológicos para auxiliar os professores na análise qualitativa das contribuições dos alunos em fóruns de discussão. O software desenvolvido, chamado Minerafórum, utiliza técnicas de mineração de texto baseadas no conceito de grafos e sua principal função é identificar se as mensagens analisadas são relevantes a um determinado tema, classificando-as com um determinado fator.

Para testar a eficiência da técnica de mineração escolhida, o autor realizou uma comparação entre resultados obtidos com o uso da ferramenta e resultados obtidos pela análise manual de um grupo de professores, os quais foram muito próximos.

A principal limitação da ferramenta proposta por Azevedo é relacionada à técnica de mineração utilizada. Conforme citado no capítulo introdutório deste trabalho, quando a análise de relevância é limitada a presença de palavras chaves, não é considerada a ordenação e estrutura frasal do texto e uma frase com um conteúdo conceitualmente incorreto pode apresentar um alto fator de relevância.

2.2 Uma aplicação para análise de opiniões com base em representação linguístico-computacional

O trabalho de Gazzola (2011) aborda o desenvolvimento de uma aplicação web capaz de identificar e classificar opiniões em textos extraídos de um ambiente virtual de aprendizagem.

A técnica utilizada é baseada na polaridade de adjetivos e se propõe – de forma automatizada - a identificar a orientação positiva ou negativa de opiniões em trechos de frases inseridas pelos alunos no ambiente virtual. A análise é realizada a partir de uma base de regras linguísticas predefinidas de acordo com a classe gramatical de cada palavra e as possíveis classificações são: Positivo, Negativo, Neutro e Não classificado.

A intenção do estudo foi criar uma ferramenta para facilitar a avaliação da satisfação dos alunos em relação ao curso. Em relação às limitações, o autor do estudo conseguiu minimizar um provável problema utilizando um conjunto de regras para definir sentenças qualificadoras, desta forma é analisada a presença de palavras na frase que podem alterar a polaridade de um adjetivo.

2.3 Template-based Question Answering over RDF Data

Artigo escrito por Unger et al. (2012) com o objetivo de apresentar uma nova abordagem estrutural para sistemas automatizados de respostas de perguntas baseados na linguagem de representação de informação RDF.

A abordagem consiste na análise da questão e construção de um modelo com a linguagem de consulta SPARQL, que representa o espelhamento da estrutura interna da pergunta. Este modelo é desenvolvido com detecção de predicado e identificação da entidade estatística. A motivação para o estudo ocorreu devido ao formato da maior parte dos sistemas de respostas de perguntas não conseguir representar fielmente a estrutura semântica da linguagem natural das questões, impossibilitando a apresentação de respostas assertivas.

O trabalho conclui que o formato é competitivo apresentando exemplos de questões que o modelo proposto consegue responder enquanto outras abordagens de sistemas concorrentes não conseguem.

2.4 Comparação

Tanto o trabalho de Azevedo quanto o de Gazzola, assim como o presente estudo, foram realizados no âmbito de ambientes virtuais de aprendizagem. O primeiro utiliza técnicas de mineração para identificação de palavras-chave, o que pode gerar situações indesejadas em determinados cenários, conforme citado no capítulo introdutório. O segundo utiliza regras e padrões linguísticos para a realização do processamento do texto em linguagem natural, prática que também será adotada nessa abordagem.

No trabalho de Unger, a utilização da linguagem SPARQL e a proposta da construção de modelos de consultas são os itens em comum a proposta de arquitetura deste trabalho. Entretanto, não existe neste trabalho o foco na análise de respostas em linguagem natural, mas sim na geração de respostas a partir da leitura de perguntas em linguagem natural.

3 ABORDAGEM PROPOSTA

A abordagem proposta neste artigo é baseada em uma estrutura modular que utilizará componentes existentes, no intuito de aproveitar sistemas já testados e reconhecidos no meio acadêmico. Na figura 1 é apresentado o fluxo e os principais componentes propostos e a seguir são descritas as características de cada etapa.

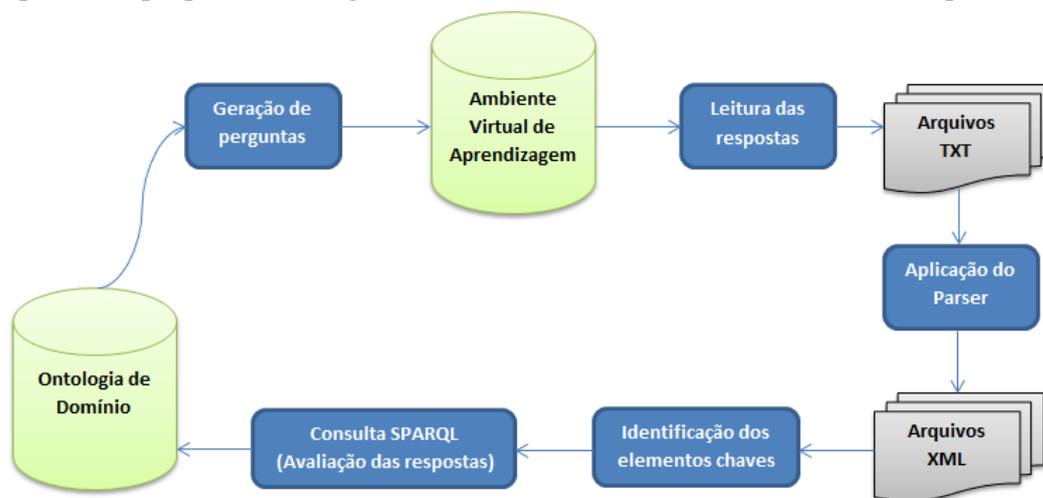


Figura 1 – Arquitetura da solução

Considera-se nesta abordagem o seguinte contexto: Uma ontologia de domínio descreve o conhecimento a respeito de uma determinada área. Esta ontologia pode ser usada, a partir de suas relações e de procedimentos baseados em PLN e geração de frases, para a geração de perguntas que podem ter suas respostas posteriormente verificadas na ontologia, através de consultas. A utilização destas perguntas em um Ambiente Virtual de Aprendizagem irá gerar respostas que serão coletadas e processadas, com uso de um parser morfofossintático. Estes resultados serão tratados com regras linguísticas para identificar os elementos importantes das respostas, que serão utilizados para a montagem de consultas em linguagem SPARQL, que possibilitam, por fim, confirmar a correção da resposta em uma consulta à ontologia de domínio.

No presente trabalho foram implementadas parcialmente estas etapas, em função das limitações de tempo existentes, tendo sido implementadas as etapas de

processamento das respostas e consultas na ontologia. A seguir são detalhadas as etapas da solução proposta.

3.1 Ontologia de domínio

A abordagem de arquitetura proposta neste artigo se inicia com a escolha da ontologia que será utilizada. Em um cenário acadêmico, educadores podem buscar por estruturas já formadas e disponíveis na web de sua área de conhecimento específica. Por exemplo, um professor do curso de Sistemas de Informação pode buscar e utilizar uma ontologia já criada que reúne os conceitos básicos de Banco de Dados.

A figura 2 apresenta uma visualização gráfica parcial de uma ontologia de domínio, reunindo alguns conceitos e relacionamentos a respeito do domínio “Banco de Dados”. Os relacionamentos “possui”, “pode ser” e “é uma” associam diferentes conceitos e possibilitam a construção de determinadas assertivas, tais como: “Um banco de dados pode ser plano”; “Um banco de dados possui uma linguagem de consulta”; “SQL é uma linguagem de consulta”. Estas assertivas podem ser utilizadas como base para a geração de perguntas em linguagem natural, com o tratamento apropriado e com o conhecimento prévio das relações desta ontologia.



Figura 2 – Visualização gráfica parcial de uma ontologia

Cada vez mais são observadas iniciativas para disponibilizar informação em formato estruturado para uso amplo. Um exemplo importante é ao projeto DBPedia (pt.dbpedia.org) que disponibiliza em formato estruturado as informações e conceitos da biblioteca Wikipedia (pt.wikipedia.org).

3.2 Geração de perguntas

Após a definição da ontologia que será utilizada como base, as perguntas devem ser criadas. Como o objetivo da solução proposta é permitir a avaliação automatizada da qualidade de respostas com fundamentação em uma base conhecimento de domínio, é necessário que a criação das perguntas seja feita de acordo com esta mesma base. Com base na ontologia apresentada na figura 2, utilizando-se a relação “possui” existente entre os conceitos “Banco de Dados” e “Linguagem de Consulta”, procedimentos heurísticos baseados em PLN podem ser utilizados para gerar, como exemplo, a seguinte pergunta: “Qual é a linguagem de consulta de um banco de dados?”.

Neste estudo utilizaremos a geração manual de perguntas após a verificação e análise visual da ontologia escolhida. Entretanto, é possível o desenvolvimento de um processo de automatização desta fase, que analisa os relacionamentos existentes na ontologia e gera perguntas automaticamente, usando técnicas de geração de linguagem natural (INDURKHAYA e DAMERAU, 2011). O desenvolvimento de tal processo será sugerido como continuação deste trabalho.

3.3 Leitura das respostas

O *corpus* considerado para a proposição desta solução é composto pelas respostas de alunos, em linguagem natural, postadas em fóruns ou outras ferramentas de interação em Ambientes Virtuais de Aprendizagem. Como o foco principal deste estudo é a realização da análise das frases, não será abordado o processo de extração destes dados dos seus ambientes e nem o seu pré-processamento.

3.4 Aplicação do Parser

Com os textos disponíveis, a próxima etapa é a análise morfossintática, ou seja, o apontamento das classes gramaticais e demais características linguísticas de cada palavra. Para isso será utilizado o parser morfossintático Palavras (BICK, 2000), pois trata-se de uma ferramenta com resultados consistentes e amplamente utilizada para o tratamento da língua portuguesa. O resultado deste processamento é a geração de um arquivo XML estruturado que contém cada sentença do texto dividida por palavras com os seus respectivos atributos.

3.5 Identificação dos elementos chaves

Nesta etapa é realizada a consulta do arquivo XML gerado pela ferramenta de parser morfossintático utilizada. O objetivo principal é aplicar regras e padrões linguísticos para extrair os elementos chaves das respostas, que serão utilizados no módulo seguinte.

Este componente da solução está baseado na participação de uma especialista em Linguística, que prestou o apoio no estudo de padrões linguísticos que podem ser observados em frases curtas contendo perguntas e respostas. Estes padrões linguísticos relacionam os componentes das frases e possibilitam a identificação dos elementos centrais que compõe as questões e as respostas para as questões. Existem diversas abordagens para a identificação de elementos de frases, sendo que algumas baseadas em padrões de palavras, na distância entre palavras ou no uso de recursos sintáticos, como a proposta aqui descrita.

A utilização do parser morfossintático, tal como apresentado no item anterior, disponibiliza uma riqueza de informações importante para ampliar a possibilidade de identificação correta dos elementos das frases. A estrutura XML resultante da classificação gramatical das palavras permite a aplicação de regras linguísticas que possuem mais flexibilidade para relacionar os elementos das frases, pois, entre outros aspectos, permitem identificar as estruturas internas que compõe as frases.

Por exemplo, esta abordagem permite a realização das análises a seguir. Para uma frase como “Quais são as principais cláusulas da linguagem SQL?” podem ser identificados os seguintes elementos: Pronome interrogativo + verbo [ser] + artigos + adjetivo + substantivo [cláusulas] + preposição + artigo + substantivo [linguagem SQL]. O conhecimento linguístico permite a identificação de elementos opcionais, tais como o elemento adjetivo, neste exemplo. Esta informação permitiria que a seguinte frase também fosse identificada corretamente: “Quais são as cláusulas da linguagem SQL?”. O mesmo ocorre com formas de escrita tais como voz ativa e voz passiva, entre outros exemplos.

3.6 Consulta SPARQL para avaliação das respostas

Com o conhecimento da ontologia de domínio, da pergunta e do elemento chave da resposta a ser analisado, o último passo é a realização da consulta na ontologia de domínio para avaliar a sua correção. As consultas são construídas utilizando a tecnologia SPARQL, que consiste em uma linguagem de consulta capaz de acessar e manipular dados de uma ontologia.

Foi utilizado um padrão de consulta, customizado a partir dos elementos identificados pelo item anterior. Desta forma, após a identificação dos termos de interesse nas respostas, são montadas consultas no formato SPARQL que permitem verificar na ontologia se os elementos identificados realmente são relacionados com os assuntos sendo questionados.

4 RESULTADOS

Com base na proposta de solução apresentada na seção anterior, foram realizados testes para avaliação para os elementos de processamento das frases no parser morfossintático, para a identificação dos elementos das respostas e para a geração da consulta SPARQL. Foi estabelecido um cenário para testes de validação preliminar da solução, onde foi realizada a construção de uma ontologia de domínio, a criação de uma pergunta e a simulação de três possíveis respostas em linguagem natural. Em seguida, foi executada a aplicação do parser morfossintático e das regras linguísticas utilizadas, para identificação dos elementos de cada resposta. Por último, foram criadas as consultas conforme um padrão SPARQL para avaliar cada resposta. Nesta seção é apresentado um exemplo completo do fluxo proposto.

4.1 Avaliação

Conforme indicado na descrição da solução proposta, a primeira etapa do fluxo é a escolha da ontologia de domínio. No exemplo que será apresentado foi criada uma ontologia com conceitos de Banco de Dados, utilizando o software Protege (<http://protege.stanford.edu>), seguindo padrões conhecidos de construção de ontologias (NOY e MCGUINNESS, 2001). Na figura 3 pode-se observar sua visualização gráfica.

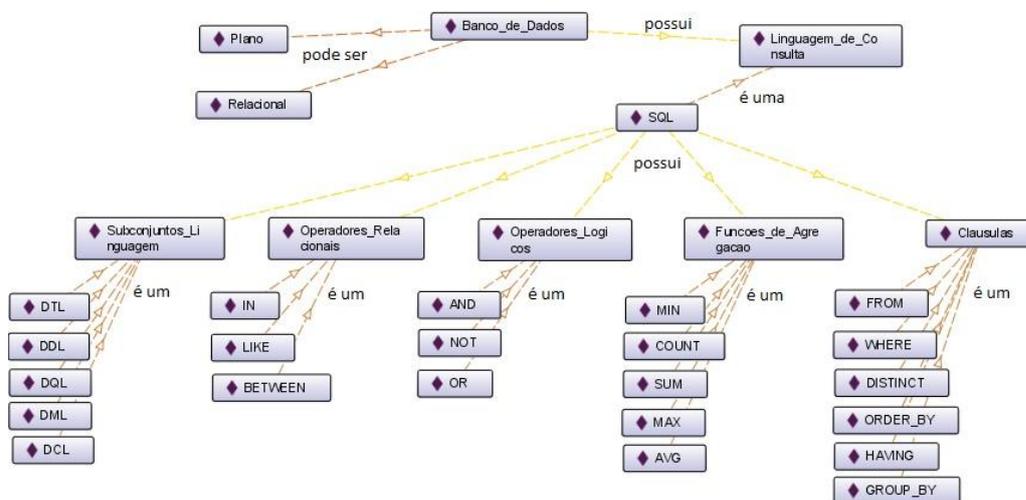


Figura 3 – Visualização gráfica da ontologia

Pode-se observar que a ontologia é formada por conceitos e relacionamentos. Como exemplo, temos o conceito “SQL” e cinco outros conceitos relacionados (Cláusulas, Funções de agregação, Operadores lógicos, Operadores relacionais e Subconjuntos de Linguagem) ligados pela relação “possui”. A partir deste cenário, pode-se definir perguntas e gerar consultas relacionadas a estes relacionamentos, tais como: O que o SQL possui? O SQL possui Cláusulas?

Para a avaliação deste estudo, foi definida uma pergunta e simuladas três possíveis respostas de alunos, conforme o quadro 1.

Pergunta:	“Quais são as principais cláusulas da linguagem SQL?”
Resposta 1:	“As principais cláusulas do SQL são: From, Where, Select e Having.”
Resposta 2:	“Order By, Group by, From, Where, Distinct e Having.”
Resposta 3:	“As principais são Select e Update.”

Quadro 1 – Respostas simuladas

Após a geração de arquivos de texto com as respostas, foi realizada a aplicação do parser morfosintático nestas frases. O quadro 2 apresenta uma parte da estrutura do arquivo XML original gerado pela ferramenta Palavras (BICK, 2000) para a resposta 1.

```
<?xml version="1.0" encoding="UTF-8"?><xml><corpus> <body>
<s id="s1" ref="1" source="Running text" forest="1" text="As principais cláusulas do SQL são:">
  <graph root="s1_500">
    <terminals>
      <t id="s1_1" word="As" lemma="o" pos="art" morph="F P" extra="artd"/>
      <t id="s1_2" word="principais" lemma="principal" pos="adj" morph="F P" extra="SUP"/>
      <t id="s1_3" word="cláusulas" lemma="cláusula" pos="n" morph="F P" extra="cc-r"/>
      <t id="s1_4" word="de" lemma="de" pos="prp" morph="--" extra="sam-"/>
      <t id="s1_5" word="o" lemma="o" pos="art" morph="M S" extra="-sam artd"/>
      <t id="s1_6" word="SQL" lemma="SQL" pos="prop" morph="--" extra="--"/>
      <t id="s1_7" word="são" lemma="ser" pos="v-fin" morph="PR 3P IND VFIN" extra="fmc mv"/>
      <t id="s1_8" word=":" lemma="--" pos="pu" morph="--" extra="--"/>
    </terminals>
  </graph>
</s>
</corpus>
</body>
</xml>
```

Quadro 2 – Estrutura parcial do arquivo XML gerado pelo parser

Com a geração dos arquivos XML, foi realizada a consulta e aplicadas as regras e padrões linguísticos para identificação dos elementos chaves das respostas. Os padrões utilizados foram obtidos com apoio de um especialista em linguística e consistem nos seguintes formatos, descritos a seguir:

a) **Pergunta**

- Frase:** “*Quais são as principais cláusulas da linguagem SQL?*”
- Padrão:** Pronome interrogativo + verbo [ser] + artigos + adjetivo + substantivo [cláusulas] + preposição + artigo + substantivo [linguagem SQL]

b) **Resposta completa**

- Frase:** “*As principais cláusulas do SQL são: From, Where, Select e Having.*”
- Padrão:** Artigo + adjetivo + substantivo [cláusulas] + preposição + artigo + substantivo [SQL] + verbo [ser] + substantivo [From, Where, Select e Having]

c) **Resposta alternativa**

- Frase:** “*Order By, Group by, From, Where, Distinct e Having.*”

- b. **Padrão:** substantivo [Order By, Group by, From, Where, Distinct] + conjunção + substantivo [Having]
- d) **Resposta padrão**
- a. **Frase:** “As principais são Select e Update.”
- b. **Padrão:** Artigo + adjetivo + verbo [ser] + substantivo [Select] + conjunção + substantivo [Update]

Como resultado da aplicação destes padrões na análise dos documentos gerados pelo parser, foram obtidas as seguintes identificações de conceitos como respostas:

Resposta 1:	“From” “Where” “Select” “Having”
Resposta 2:	“Order By” “Group by” “From” “Where” “Distinct” “Having”
Resposta 3:	“Select” “Update”

Quadro 3 – Elementos das respostas

A última etapa do fluxo proposto pela solução foi a realização das consultas SPARQL, utilizando os elementos identificados no passo anterior. Para isso empregou-se um padrão de consulta que retorna a quantidade de elementos corretos presentes em cada resposta, desta forma possibilitando a comparação com a quantidade total de itens corretos e gerando uma porcentagem de acerto, ao invés do simples conceito de certo ou errado.

A seguir é apresentado um exemplo de consulta utilizada na validação, juntamente com seu resultado. A figura 4 apresenta a consulta (realizada no software Protegé) com o retorno da quantidade total de itens corretos presentes na resposta 1. Pode ser observado na imagem que o resultado foi o número inteiro “3”, ou seja, dos quatro elementos presentes na resposta, três estão corretos.

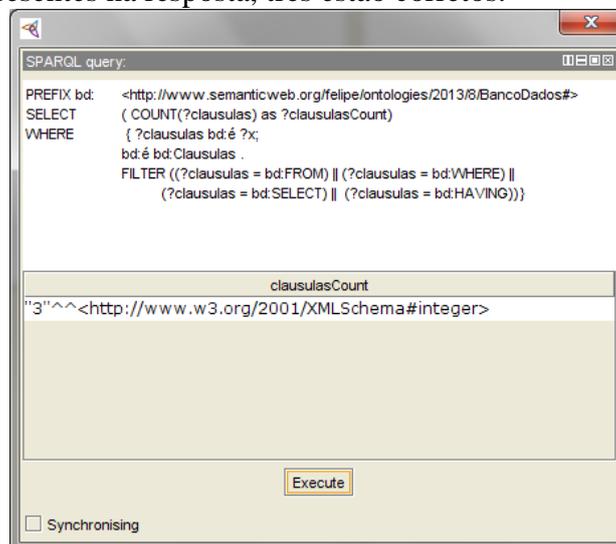


Figura 4 – Consulta e quantidade total de itens corretos presentes na resposta 1

Esta mesma consulta é realizada para cada uma das outras respostas e também é realizada diretamente na ontologia, para identificar quantidade total de itens (cláusulas) corretos, ou seja, quantas cláusulas, ao todo, o SQL possui. Com estes retornos, é realizada uma classificação dos elementos corretos.

Na tabela 1, pode-se observar o retorno das três consultas realizadas e como eles são utilizados para gerar a avaliação final. Os mesmos resultados podem ser obtidos de forma automática, através de um protótipo implementado na linguagem Java e com a API Jena para validar a proposta descrita.

Resposta	Quantidade de itens corretos	Quantidade total	Avaliação
Resposta 1	3	6	50%
Resposta 2	6	6	100%
Resposta 3	0	6	0%

Tabela 1 - Avaliação das respostas

5 CONCLUSÃO

Diante de um cenário onde o professor tem o papel de analisar um volume muito alto de informações para realizar suas avaliações e interagir com os alunos, buscou-se o uso de recursos tecnológicos para agrupar e transformar estes dados em informações relevantes para o apoio à tomada de decisões. O presente artigo estudou possibilidades de auxílio à avaliação de textos em ambientes de aprendizado virtuais a partir do emprego de tecnologias de processamento de linguagem natural e ontologias.

O principal objetivo deste trabalho foi concluído, considerando que foi proposta e através de um protótipo implementado foi avaliada empiricamente uma abordagem de solução baseada na utilização de ontologias e PLN que serve de apoio na avaliação de textos de respostas provenientes de interações em um ambiente virtual de aprendizagem.

REFERÊNCIAS

- AZEVEDO, Breno F. T. **MINERAFÓRUM: Um recurso de apoio para análise qualitativa em fóruns de discussão**. PGIE-UFRGS, Porto Alegre, 2011.
- BICK, E. “**The Parsing System PALAVRAS: Automatic Gramatical Analysis of Portuguese in a Constraint Grammar Framework**”. Århus University. Århus: Århus University Press. 2000.
- BLOOM, Benjamin S.; CARROLL, John Bissell. **Mastery learning: Theory and practice**. Ed. James H. Block. New York: Holt, Rinehart and Winston, 1971.
- GAZZOLA, Oliver. **Uma aplicação para análise de opiniões com base em representação linguístico-computacional**. UNISINOS, São Leopoldo, 2011.
- GRUBER, T. R. **What is an Ontology?** Disponível em: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>, 1992. Acesso em: Maio de 2013.
- INDURKHYA N., DAMERAU, F. J., **Handbook of natural language processing**, CRC Press. New York. 2011
- LIDDY, R. **Natural Language Processing, Library and Information Science**, Marcel Drexler Inc. New York, USA, 2a Ed. 2003.
- NOY, F. N.; MCGUINNESS, D. L. **Ontology development 101: a guide to create your first ontology**. Disponível em: <http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.doc>. Acesso em: Maio de 2013.
- UNGER, Christina et al. **Template-based Question Answering over RDF Data**. WWW 2012 – Session: Ontology Representation and Querying: RDF and SPARQL April, 2012, Lyon, France.