

# Sistema multiagentes para recuperação e análise de informações como suporte à mediação pedagógica em ambientes colaborativos de aprendizagem

Sabrina Siqueira Panceri<sup>1,2</sup>, Crediné Silva de Menezes<sup>1,3</sup>

<sup>1</sup>Programa de Pós-Graduação em Informática – Universidade Federal do Espírito Santo  
Av. Fernando Ferrari, 514, Goiabeiras – Vitória – ES – Brasil – Caixa Postal 29075-910

<sup>2</sup>Campus Guarapari – Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo  
Estrada da Tartaruga, s/n, Muquiçaba – Guarapari – ES – Brasil – Caixa Postal 29215-090

<sup>3</sup>Faculdade de Educação – Universidade Federal do Rio Grande do Sul  
Av. Paulo Gama, s/n, Farroupilha – Porto Alegre – RS – Brasil – Caixa Postal 90046-900

sabrina.panceri@ifes.edu.br, credine@gmail.com

**Abstract.** *This paper presents a proposal of a Multi-Agent System, which aims to recover and analyze information through the mining of texts produced in virtual teaching and learning environments. The Educational Architecture Thesis Debate presented as favorable scenario the application of SMA. They are elicited pedagogical mediations within the context of a Thesis Debate. The development of functional prototype of SMA and its application to a specific case are described. Pedagogical mediations that are aided by the use of SMA are teased. Finally, the results achieved in the final working considerations are exposed.*

**Resumo.** *Este trabalho apresenta uma proposta de um Sistema Multiagente (SMA), que visa a recuperação e análise de informações, através da mineração dos textos produzidos em ambientes virtuais de ensino e aprendizagem. A Arquitetura Pedagógica Debate de Teses apresenta-se como cenário favorável a aplicação do SMA. São elicítadas as mediações pedagógicas dentro do contexto de um Debate de Teses. O desenvolvimento do protótipo funcional do SMA e sua aplicação a um caso específico são descritos. As mediações pedagógicas que são auxiliadas pelo uso do SMA são descritas com detalhes. Por fim, os resultados alcançados e as considerações finais do trabalho são expostos.*

## 1. Introdução

Diante dos avanços tecnológicos ocorridos na última década, a internet se apresenta como parte integrante do cotidiano da atual sociedade. Dessa forma, as diversas formas de comunicação adaptaram-se para sua integração e utilização como meio para propagação e compartilhamento de informações. Assim também as práticas educacionais incorporaram a internet como meio para promovê-las, difundindo cada vez mais a educação à distância com base nesse novo aporte computacional.

Com o objetivo de oferecer suporte à educação a distância, diversas soluções computacionais foram desenvolvidas, dentre elas destacam-se os Ambientes Virtuais de Ensino e Aprendizagem (AVEA), que são sistemas colaborativos concebidos para

favorecer a mediação de aprendizagens. Nos AVEA atuais, observa-se a presença de variadas ferramentas de autoria colaborativa, tais como blogs, *wikis* e fóruns.

As interações praticadas nas ferramentas de autoria são, em grande maioria, feitas através da produção de textos, seja na elaboração de uma resposta a uma pergunta, nas postagens realizadas nos fóruns, ou mesmo na produção de uma redação de forma conjunta com outros participantes. Este volume de textos produzidos subsidia um nicho de pesquisa ainda pouco explorado, a Recuperação de Informações em Textos produzidos dentro do contexto educacional.

Diante desse cenário, propõe-se o desenvolvimento de um Sistema Multiagente (SMA) que servirá como uma camada de análise e que poderá ser acessado por qualquer AVEA por meio de interfaces comunicação predefinidas. O SMA é responsável por processar, por intermédio de seus agentes, os textos produzidos nas ferramentas de autoria, gerando seis tipos de resultados. Os resultados gerados pelo SMA servem para apoiar decisões e mediações pedagógicas.

Este artigo está organizado da seguinte maneira: Na Seção 2 apresenta-se brevemente a fundamentação teórica utilizada para a construção da proposta de solução que é detalhada na Seção 3. Na Seção 4 mostra-se a avaliação da proposta e sua aplicação a um caso real, bem como a análise dos resultados obtidos. Na Seção 5 estão os trabalhos correlatos ao contexto e proposta aqui apresentados e por fim, na Seção 6 estão as considerações finais deste trabalho.

## **2. Fundamentação Teórica**

Esta seção apresenta os fundamentos teóricos utilizados para a construção de uma solução computacional inteligente, baseada na arquitetura de Sistemas Multiagente, para dar suporte tecnológico à aplicação de Arquiteturas Pedagógicas, especificamente, para aplicação da Arquitetura Pedagógica Debate de Teses. Bem como um breve resumo sobre a Recuperação de Informações em Textos que dá suporte a construção do núcleo de processamento dos textos que serão analisados pela solução proposta.

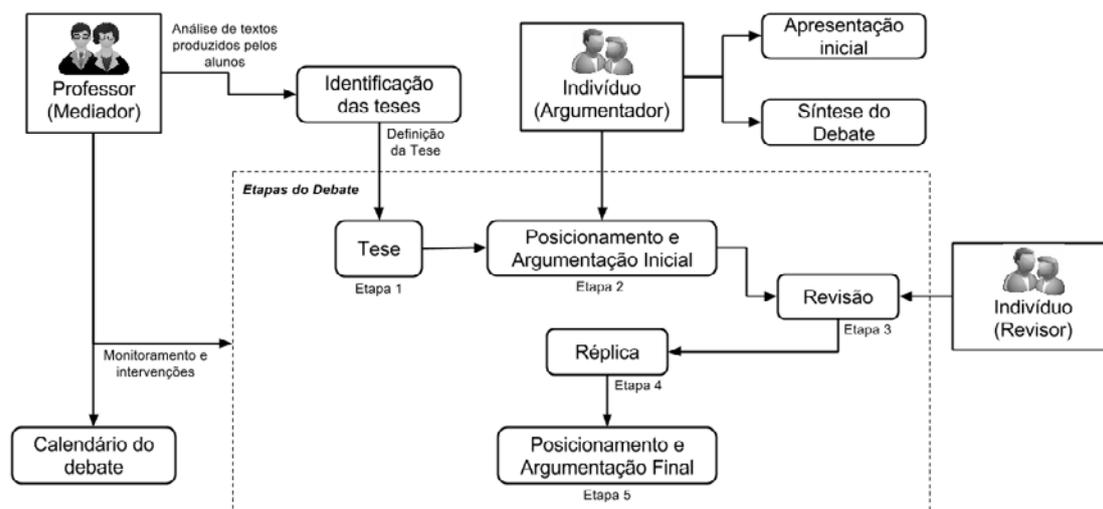
### **2.1. Arquiteturas Pedagógicas**

Para subsidiar a formulação de propostas de uso das tecnologias digitais na educação, em [Carvalho et al. 2005] é apresentado o conceito de Arquiteturas Pedagógicas, que destinam-se a aplicação de abordagens pedagógicas baseadas na Epistemologia Genética de Jean Piaget [Ramoszi-Chiarottino 1998].

A proposta dos autores [Carvalho et al. 2005] tem por objetivo organizar as interações em ambientes virtuais de aprendizagem, adotando protocolos específicos, com a finalidade de auxiliar a construção de conhecimento dos indivíduos de forma colaborativa.

#### **3.1.1 A Arquitetura Pedagógica Debate de Teses**

Uma especialização das Arquiteturas Pedagógicas é apresentada em [Nevado et al. 2011] com a designação de Arquitetura Pedagógica Debate de Teses (APDT). A APDT propõe a construção de conhecimento de forma colaborativa por meio de interações orientadas. Uma visão geral sobre o fluxo de interações realizadas na APDT é apresentado pela Figura 1.



**Figura 1: Fases de um Debate de Teses**

Na APDT (Figura 1) o professor, no papel de mediador, pela definição do calendário (prazos) de cada etapa do debate, pelo monitoramento das interações realizadas pelos alunos, e pela Definição da Tese (Etapa 1) através da análise de textos produzidos anteriormente sobre um determinado tema. Uma tese é uma afirmação sobre qualquer assunto, sobre a qual o aluno pode concordar ou discordar. Finalizada as definições iniciais, o professor monitora as interações e realiza, se necessário, pequenas intervenções a fim de auxiliar os alunos a desenvolver da melhor maneira os textos de cada etapa do debate.

Os alunos são responsáveis pelas produções das demais etapas da APDT, iniciando sua participação no debate através da indicação de seu Posicionamento e Argumentação Inicial sobre a tese (Etapa 2). O posicionamento deve ser favorável ou contrário à tese em discussão, e o argumento deve ser desenvolvido de forma lógica e com base nos conhecimentos prévios e que justifique o posicionamento indicado.

A Revisão (Etapa 3) é realizada por outros dois alunos – revisão por pares. Estes alunos devem analisar a consistência dos argumentos, tendo por base verificar se há evidências que sustentem a argumentação em revisão. O aluno na etapa de revisão assume o papel de revisor.

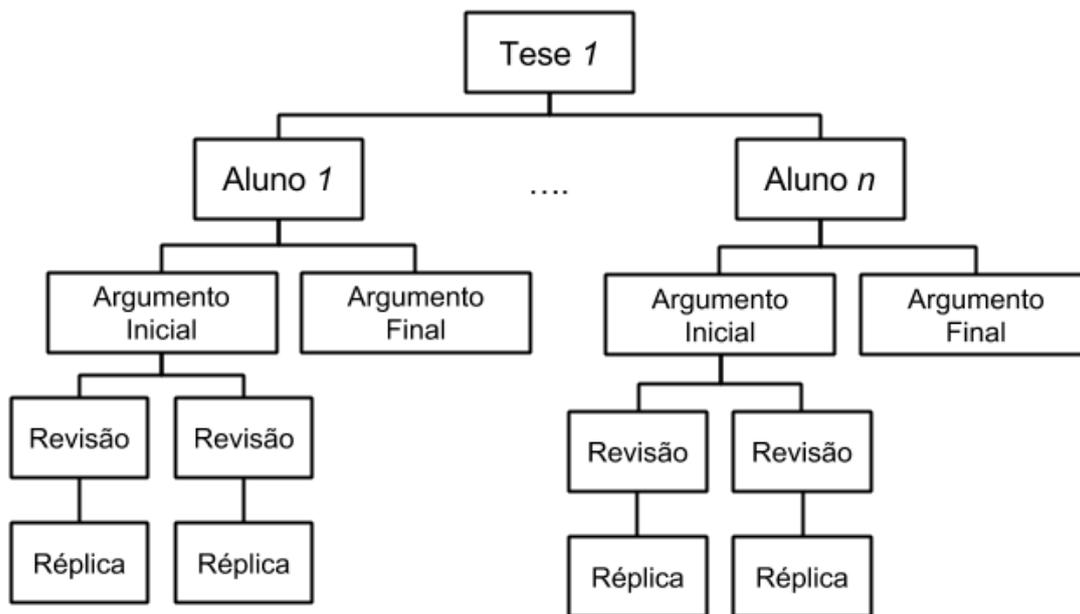
Na etapa de réplica (Etapa 4), o aluno volta ao papel de argumentador, e deve responder aos levantamentos feitos por seus revisores. As Etapas 3 e 4 preparam o aluno para a reescrita de seu posicionamento e argumentação, que é concretizado na Etapa 5 – Posicionamento e Argumentação Final.

Nesta última etapa, o aluno indica seu posicionamento sobre a tese e reescreve sua argumentação, levando em conta as interações realizadas. Ao final da aplicação da APDT, espera-se que o aluno tenha consolidado seus conhecimentos prévios sobre o assunto debatido.

As interações nesta Arquitetura Pedagógica ocorrem mediante a autoria de textos, geralmente, textos curtos. Para que sejam aplicadas mediações pedagógicas pertinentes a cada etapa, o professor (mediador) deve ler e compreender todos os textos produzidos. Considerando uma turma com dez (10) alunos, e seguindo a Figura 1, para

cada tese proposta o professor terá dez (10) posicionamentos e argumentações iniciais, vinte (20) revisões, vinte (20) réplicas, e dez (10) posicionamentos e argumentações finais, num total de sessenta (60) textos para analisar, compreender e identificar as características intrínsecas de cada etapa.

Uma abordagem possível para dar suporte computacional a Arquitetura Pedagógica Debate de Teses é a configuração de um fórum com regras específicas. Ao desenvolver a APDT utilizando a ferramenta fórum como suporte, cada tese pode ser definida na forma de um tópico do fórum (etapa 1) e as respostas realizadas por cada aluno ao tópico equivalem a etapa 2 de posicionamento e argumentação inicial. Nesta resposta, o aluno deve manifestar sua concordância ou discordância sobre a afirmação proposta, e em seguida fazer sua argumentação de forma lógica. Para cada resposta enviada, são designados dois revisores, que apresentam problematizações ao autor da resposta, caracterizando assim a etapa 3 – revisão. Os argumentadores podem responder a cada uma das revisões (etapa 4) e finalmente devem produzir uma nova resposta à tese (etapa 5 – posicionamento e argumentação final). Uma visualização dessa abordagem está na Figura 2.



**Figura 2: Árvore de interações para produção de um Debate de Teses utilizando a ferramenta Fórum**

Contudo, a prática da APDT com o uso da ferramenta fórum fica prejudicada uma vez que a organização dos textos, o acesso as produções e demais interações ficam encadeadas ao tópico proposto, dificultando o monitoramento e possíveis intervenções do professor. Diante disso, um ambiente próprio foi desenvolvido para dar suporte a aplicação da APDT. Neste ambiente<sup>1</sup> o fluxo de interações segue a Figura 1. O objetivo da presente pesquisa é apresentar uma colaboração ao ambiente em funcionamento, voltada para apoiar o professor nas tarefas subsequentes a realização do debate, fornecendo suporte computacional às tarefas de análise dos textos para subsidiar a

<sup>1</sup> O ambiente pode ser acesso em <http://www.pead.faced.ufrgs.br/sites/cms/debate/>. Acesso em 25 de março de 2016.

aplicação das mediações pedagógicas.

## 2.2. Recuperação de Informações em Textos

A Recuperação de Informações (RI) tem como foco a busca de informações em documentos ou textos, e a classificação destas buscas a fim de retornar os documentos mais importantes em uma coleção, dada uma quantidade limitada de informações a serem resgatadas. De acordo com [Manning et al. 2008], “RI busca encontrar material de natureza não estruturada que satisfaçam uma necessidade de informação, a partir de grandes coleções”.

Outro importante aspecto das pesquisas em RI é a utilização do Modelo Espaço Vetorial para representação computacional das coleções que serão analisadas [Salton et al. 1975]. Este modelo representa os documentos de uma coleção como vetores em um espaço multidimensional. Para calcular a frequência ou ocorrência dos termos que compõem os documentos da coleção, aplica-se a medida TF-IDF - *Term Frequency x Inverse Document Frequency*. Nesta medida, a importância de um termo é proporcional à sua frequência de ocorrência em cada documento da coleção, e inversamente proporcional ao número de documentos em que o termo aparece. Esse cálculo é feito de acordo com a Fórmula 1:

$$TF-IDF(w, d) = TermFreq(w, d) \cdot \log(N / DocFreq(w)) \quad (1)$$

Onde  $TF-IDF(w, d)$  é a frequência de cada palavra no documento,  $N$  é o número de todos os documentos da coleção, e  $DocFreq(w)$  é o número de documentos que contém o termo  $w$ .

Dessarte, o vetor que representa cada documento contém todos os termos utilizados em todos os documentos da coleção. Sendo assim, quanto maior a coleção, maior a dimensionalidade do espaço vetorial que a representa. Para isso, a aplicação de técnicas de pré-processamento na coleção que será analisada pode melhorar a qualidade de representação vetorial da mesma, uma vez que há a redução nos termos que compõem a coleção. No contexto deste trabalho, utilizou-se as seguintes técnicas:

- Remoção de termos frequentes (*stopwords*): Essa técnica visa a remoção de termos comuns, geralmente pertencentes as classes gramaticais dos artigos ou preposições e de palavras pequenas (com menos de 3 caracteres). Essas palavras, geralmente, não possuem relevância semântica e não interferem na recuperação de informações nos textos.
- Capitalização (*case-folding*): Essa técnica troca os caracteres escritos em maiúsculo para minúsculo, para que palavras como por exemplo “Interessante” e “interessante” sejam consideradas as mesmas.
- Remoção de caracteres acentuados: A troca de caracteres acentuados (ã, é, õ) pelo seu correspondente sem acentuação (a, e, o) auxilia no processo de análise dos termos e na padronização dos termos que serão analisados.
- Redução ao radical (*stemming*): O objetivo do *stemming* é reduzir palavras que se encontram em formas derivadas para o seu radical de formação (*stem*), através da remoção dos afixos (prefixos ou sufixos).

O modelo de representação do corpus é o “*bag-of-words*”, no qual todas as

palavras que compõem o documento são consideradas suas características. Este modelo não considera a ordem em que as palavras aparecem no texto.

Segundo [Manning et al. 2008] a Clusterização é a forma mais comum de aprendizado não supervisionado, ou seja, não há interferência humana no processo. Utiliza-se algoritmo de clusterização *K-Means* para particionar o corpus em um conjunto pré-definido de grupos de semelhança (*clusters*). Os agrupamentos são feitos com base na semelhança encontrada entre os textos. Por padrão o *K-Means* utilizada a Distância Euclidiana para calcular a semelhança entre os textos, porém [Feldman e Sanger 2006] afirma que a Similaridade de Cossenos (*Cosine Similarity*) é a medida de similaridade mais utilizada para realizar o agrupamento de textos, dessa forma, utiliza-se neste trabalho uma adaptação do *K-Means*, na qual a Similaridade de Cossenos é empregada para calcular a similaridade dos textos, e assim criar os grupos de semelhança.

Para criar os dicionários de contexto e relações semânticas, utiliza-se a Análise Semântica Latente (LSA, do inglês *Latent Semantic Analysis*) [Dumais 2004]. A LSA é uma técnica estatística que visa a construção de matrizes de relações entre os documentos e seus termos, considerando que termos que estão num mesmo documento podem conter determinada relação semântica. Apesar de considerar a co-ocorrência dos termos num documento, LSA utiliza o modelo de representação “*bag-of-words*”, e não leva em consideração a sequência em que os termos aparecem no documento. Para que a aplicação do LSA seja bem-sucedida, é necessário que seja feito o treinamento da base que será usada para a análise de similaridade dos textos.

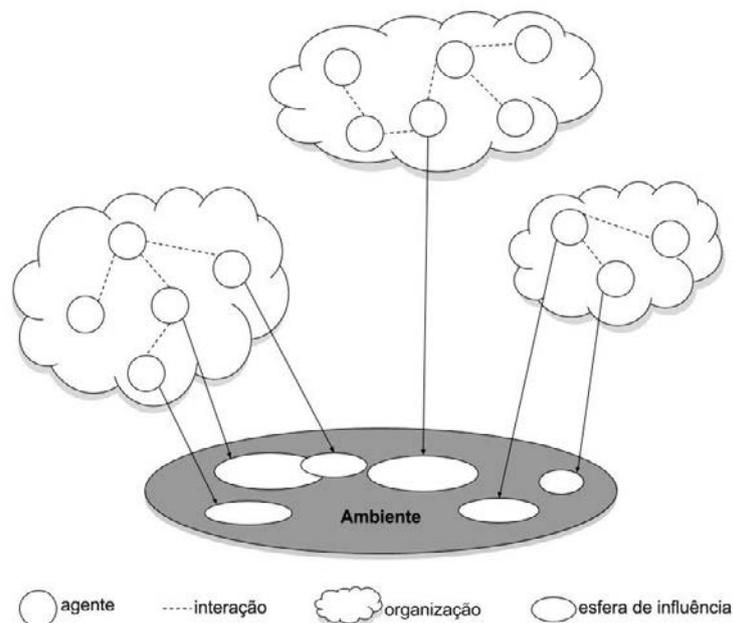
### 2.3. Sistemas Multiagente

De acordo com [Reis 2003], Sistemas Multiagente (SMA) são “sistemas computacionais em que dois ou mais agentes interagem ou trabalham em conjunto de forma a desempenhar determinadas tarefas ou satisfazer um conjunto de objetivos”.

Segundo [Wooldridge 2009] um agente é “um sistema de computador que está situado em algum ambiente, e que é capaz de ação autônoma neste ambiente, a fim de cumprir os objetivos para os quais fora projetado”.

Agentes inteligentes podem ser classificados em dois tipos: Agentes Inteligentes Reativos e Agentes Inteligentes Cognitivos. Os Agentes Inteligentes Reativos são aqueles que reagem mediante a condição percebida em um cenário real ou fictício. Os Agentes Inteligentes Cognitivos são aqueles que consideram uma série de fatores para realizar uma reação complexa e bem elaborada. Além disso, um agente inteligente deve possuir atributos que o possibilitem interagir com o meio, são eles os atributos sensores e atributos atuadores. Atributos sensores permitem que os agentes inteligentes capturem informações do meio sobre sua influência. Atributos atuadores possibilitam a aplicação das reações dos agentes inteligentes.

Os agentes inteligentes que compõem um SMA possuem uma área de atuação ou influência, que pode ser compartilhada com outros agentes inteligentes, e que fazem parte do SMA como um todo, como mostra a Figura 3



**Figura 3: Estrutura típica de um Sistema Multiagente [Jennings 2000]**

Segundo [Coppin 2004] em muitas situações, um SMA pode ser composto apenas por agentes reativos, uma vez os agentes não existem de forma isolada e na prática, os agentes são compostos por um conjunto de regras, logo suas ações são limitadas. Ainda assim, SMA são considerados sistemas complexos, tanto para sua modelagem, quanto para seu desenvolvimento.

### 3. Contexto do Problema

No contexto dos cursos de graduação ou pós-graduação realizados na modalidade a distância, as interações realizadas entre professores e alunos são feitas através da troca de mensagens, sendo o fórum um dos principais instrumentos apoiadores dessa comunicação. A forma mais comum de uso da ferramenta Fórum é o suporte ao esclarecimento de dúvidas (pergunta-resposta). Contudo, outras atividades de interação podem ser promovidas através dessa ferramenta como por exemplo a discussão sobre um assunto determinado, o desenvolvimento de histórias colaborativas, o esclarecimento de dúvidas diversas sobre o conteúdo das disciplinas. Em todas essas possibilidades há uma vasta produção de pequenos textos.

Os textos produzidos nas ferramentas de autoria dos AVEA são analisados pelos professores e/ou tutores, com o objetivo de oferecer o suporte necessário para apoiar a construção de conhecimento dos alunos. A análise destes textos demandam a dedicação de tempo e atenção de seus analisadores, que segundo [Mason e Grove-Stephensen 2002] essa atividade consome cerca de 30% do tempo de trabalho dos analisadores. Essa quantidade de tempo é expressiva e minimizá-la incentiva a criação de ferramentas computacionais que realizem a análise e recuperação de informações para auxiliar o professor neste processo de análise.

Buscando contribuir para a superação das dificuldades que um professor enfrenta

neste esforço de mediação, propõe-se o desenvolvimento de um Sistema Multiagente (SMA) que visa extrair informações relevantes através da análise dos textos produzidos. A arquitetura geral da proposta é apresentada na Figura 4. Nela o SMA é composto por seis agentes cada um deles produzindo resultados específicos.



**Figura 4: Arquitetura Geral da Proposta**

O SMA é responsável apenas pela análise dos textos, a extração dos textos contidos nas ferramentas de autoria deve ser feita de acordo com as configurações do AVEA. A interface de comunicação com o SMA é através do envio de arquivos de texto, no qual cada linha do arquivo será considerada uma produção distinta a ser analisada.

De acordo com o resultado desejado pelo usuário, os agentes são acionados e executam processos comuns à área de Recuperação de Informações em Textos. A seguir descreve-se as responsabilidades de cada agente previsto na Figura 4.

- Agente de Pré-processamento: responsável por coordenar os processos de pré-processamento textual, que são: remoção de *stopwords*; remoção de acentuação; aplicação de *stemming*; remoção de *links*;
- Agente de Sintaxe: Responsável por verificar a formação das orações que compõem os textos, classificar as palavras de acordo com sua ordem nas orações através da utilização do Projeto Floresta Sintática<sup>2</sup> [Afonso et al. 2001].
- Agente de Produção Textual: Responsável por realizar a análise de similaridade entre os textos, através do algoritmo de Similaridade de Cossenos. Utiliza do Agente de Contexto, a Análise de Semântica Latente (LSA) [Dumais 2004] para criar dicionários de contexto com base nos próprios textos que serão analisados, e a partir desses dicionários calcular as relações semânticas entre os termos e destacar os termos mais importantes da coleção.
- Agente de Clusterização: Responsável por aplicar algoritmos de clusterização, como o *K-Means*. Os grupos são definidos com base na análise de similaridade entre os textos.
- Agente de Análise textual: Responsável por identificar se os textos produzidos contém elementos novos, a fim de contribuir com a construção de conhecimentos dos outros membros do AVA.
- Agente de Contexto: Responsável por verificar se os textos produzidos são de

2 <http://www.linguatca.pt/Floresta/>

determinado tema. O tema deve ser informado através do envio de materiais didáticos para criação dos dicionários de contexto. Essa criação é realizada através da aplicação da Análise Semântica Latente (LSA).

Os agentes serão acionados conforme a solução desejada. Segue a descrição das soluções.

- Grupos de similaridade: O usuário deve fornecer a quantidade de grupos em que ele deseja dividir sua base de análise. Produto da aplicação dos Agentes de Pré-processamento e Clusterização, os resultados apresentados são grupos de textos semelhantes.
- Destaque de ideias: O Agente de Sintaxe em parceria com os agentes de Pré-Processamento, Contexto e Produção Textual, analisa os textos a fim de encontrar frases que ressaltam informações novas sobre o tema e as mostra para o usuário.
- Análise de plágio: O Agente de Produção Textual separa os textos que possuem um nível de semelhança muito alto e os apresenta ao professor, para que este o analise e confira se há uma relação de cópia/plágio entre os textos.
- Análise de contribuições: Verificar se as informações dos textos possuem relevância e apresentam novos elementos a fim de colaborar com a construção de conhecimento dos membros do AVEA. Utiliza o Agente de Sintaxe.
- Identificação de polaridade: Destaca as afirmações positivas ou negativas identificadas nos textos analisados. Essa análise utiliza os Agentes de Contexto e Produção Textual.
- Identificação de dúvidas: Verifica expressões de dúvidas, mesmo que não tenham sido escritas como uma. Utiliza os agentes de Sintaxe, Produção Textual e Contexto.

#### 4. Avaliando o SMA proposto

Para apoiar a utilização e aplicação da Arquitetura Pedagógica Debate de Teses, foi desenvolvido um ambiente virtual de aprendizagem próprio apresentado em [Nevado et al. 2011]. Com a aplicação deste protótipo ao longo dos anos, diversas mediações pedagógicas foram identificadas, contudo aplicá-las sem o devido suporte computacional mostrou-se inviável. As mediações elicitadas são [Panceri e Menezes 2014a, Panceri e Menezes 2014b]:

1. Agrupar participantes por posicionamento e argumentação inicial;
2. Fazer comentários sobre a argumentação inicial para orientar o participante;
3. Verificar se as revisões feitas pelo participante são diferentes e relevantes;
4. Verificar se não houve comparações entre argumentações revisadas;
5. Conferir se as réplicas respondem aos questionamentos feitos nas revisões;
6. Verificar se existem novos argumentos que reforcem o posicionamento;
7. Identificar os “pontos importantes” para orientar a fase de reflexão;
8. Elaborar síntese das argumentações finais para realizar o *feedback*;

9. Agrupar argumentações finais semelhantes;
10. Verificar qual o grau de influência das revisões e réplicas na argumentação final;
11. Identificar quais teses podem ser trabalhadas a partir das argumentações finais;
12. Analisar o padrão de respostas de cada participante;
13. Verificar a evolução das argumentações;
14. Avisar sobre o prazo de envio das etapas;
15. Verificar se houve plágio nas revisões, réplicas e argumentações finais;
16. Verificar se houve cordialidade nas redações.

Aplicando o SMA para auxiliar nas mediações pedagógicas identificadas, considera-se como *corpus* de análise os textos produzidos em cada fase do Debate de Teses. Dessa forma, para cada *corpus* retirado de uma das etapas do debate, teremos resultados e análises diferenciadas, vinculados às mediações pedagógicas identificadas para cada etapa do debate.

Neste artigo, o foco é a apresentação da ferramenta desenvolvida para dar suporte as mediações pedagógicas “Agrupar participantes por posicionamento e argumentação inicial” e “Agrupar argumentações finais semelhantes”. Agrupar os alunos com base na análise do posicionamento e argumentação inicial serve para facilitar a distribuição de revisores e para identificar quais alunos iniciaram o DT com o nível de conhecimentos prévios semelhantes. Dessa forma, o professor mediador poderá sugerir materiais didáticos para auxiliar nas próximas fases do DT, e indicar revisores que estejam em grupos diferentes, para estimular que as revisões sejam feitas sob pontos de vistas diferentes. No caso do agrupamento com base no posicionamento e argumentação final, o professor terá os grupos de alunos que chegaram a conclusões semelhantes sobre o assunto discutido.

#### 4.1. Instanciando o SMA para um Debate de Teses

Para comprovar a eficácia do SMA proposto, uma instância dele foi implementada e a ferramenta **Grupos de Similaridade** foi aplicada sobre os textos produzidos num debate de teses realizado no ano de 2013. O debate em questão foi realizado como parte do curso de extensão promovido na XIV Maratona de Empreendedorismo da UFRGS [Michels 2014]. As teses propostas foram:

- (1) “O mundo mudou, o perfil do consumidor mudou, e, conseqüentemente, a forma de se planejar negócios também mudou”
- (2) “O plano de negócios aumenta a probabilidade de sucesso dos negócios”
- (3) “A visão macro do negócio é desenvolvida com o auxílio do modelo de negócios e a visão micro com o auxílio do plano de negócios. Logo, o empreendedor precisa ter visão do todo e, se necessário, a visão de cada elemento do negócio”

A amostra analisada pela ferramenta é composta pelos 21 alunos que concluíram todas as etapas do debate. Os experimentos foram realizados para divisão do total de alunos em 3, 4 e 5 grupos, e os resultados dos agrupamentos são apresentados nas Tabelas 1, 2 e 3.

A Tabela 1 apresenta os resultados obtidos sobre o debate desenvolvido a partir da análise da Tese (1).

**Tabela 1. Resultados dos experimentos realizados para agrupar as argumentações iniciais produzidas sobre a Tese (1)**

<b>Divisão em 3 grupos</b>				
Grupos	G1	G2	G3	
Qtde alunos	6	9	6	
Similaridade média	75%	51%	65%	

<b>Divisão em 4 grupos</b>				
Grupos	G1	G2	G3	G4
Qtde alunos	3	9	4	5
Similaridade média	61%	49%	68%	78%

<b>Divisão em 5 grupos</b>					
Grupos	G1	G2	G3	G4	G5
Qtde alunos	3	3	4	8	3
Similaridade Média	54%	85%	80%	52%	61%

Analisando os resultados apresentados pela Tabela 1, observa-se que:

- ◆ A similaridade média entre os membros dos grupos formados com a formação de 5 grupos é maior que 50%.
- ◆ Para a formação de 4 grupos, houve uma variação maior entre a quantidade de membros entre os grupos.
- ◆ O G2 formado com a divisão do total em 3 grupos possuem uma quantidade maior de alunos e uma baixa similaridade média entre seus membros.

A Tabela 2 apresenta os resultados obtidos sobre o debate desenvolvido a partir da análise da Tese (2). Analisando os resultados apresentados pela Tabela 2, observa-se que:

- ◆ Nas três divisões realizadas, um grupo destaca-se por ter uma similaridade média entre seus membros de 92%
- ◆ Na divisão em 3 grupos, o grupo G1 possui os mesmos alunos do grupo G3 formado com divisão em 4 grupos. Ambos os grupos possuem uma similaridade média entre seus membros de 92%.
- ◆ A distribuição mais balanceada, em relação ao número de alunos por grupos, deu-se na divisão em 5 grupos de alunos.

**Tabela 2. Resultados dos experimentos realizados para agrupar as argumentações iniciais produzidas sobre a Tese (2)**

<b>Divisão em 3 grupos</b>					
Grupos	G1	G2	G3		
Qtde alunos	4	10	7		
Similaridade média	92%	82%	46%		

<b>Divisão em 4 grupos</b>					
Grupos	G1	G2	G3	G4	
Qtde alunos	9	5	4	3	
Similaridade média	81%	59%	92%	28%	

<b>Divisão em 5 grupos</b>					
Grupos	G1	G2	G3	G4	G5
Qtde alunos	3	3	6	5	4
Similaridade Média	44%	92%	85%	75%	41%

**Tabela 3. Resultados dos experimentos realizados para agrupar as argumentações iniciais produzidas sobre a Tese (3)**

<b>Divisão em 3 grupos</b>					
Grupos	G1	G2	G3		
Qtde alunos	4	12	5		
Similaridade média	82%	66%	47%		

<b>Divisão em 4 grupos</b>					
Grupos	G1	G2	G3	G4	
Qtde alunos	1	12	4	4	
Similaridade média	0%	66%	81%	58%	

<b>Divisão em 5 grupos</b>					
Grupos	G1	G2	G3	G4	G5
Qtde alunos	4	4	8	4	1
Similaridade Média	65%	81%	64%	58%	0%

A Tabela 3 apresenta os resultados obtidos sobre o debate desenvolvido a partir da análise da Tese (3). Analisando os resultados apresentados pela Tabela 3, observa-se que:

- Para a divisão em 4 ou 5 grupos, os agrupamentos apresentam uma argumentação destoante das outras, que fica isolada em grupo de 1 membro.
- Na divisão em 4 e 3 grupos, os grupos G2 contém os mesmos 12 alunos em ambas distribuições.

Sobre a aplicação da ferramenta Grupos de Similaridade, destaca-se:

(a) Os grupos formados com a aplicação do Agente de Clusterização, parte central da ferramenta Grupos de Similaridade, são suficientes para auxiliar o professor a fornecer *feedback* para os grupos, reduzindo assim o seu esforço na produção do *feedback* individual para todos os alunos;

(b) A utilização dos agrupamentos para colocar em prática os princípios Piagetianos de “desequilíbrios”, indicando revisores que não estão num mesmo grupo a fim de estimular o acesso a pontos de vistas diferentes sobre uma mesma tese, dessa forma o revisor irá contrapor os argumentos do outro aluno. Por consequente, na réplica, o aluno revisado terá acesso a contrapontos sobre sua argumentação, o que causará certo desequilíbrio, dessa forma o aluno revisado terá que buscar por novos conhecimentos para conseguir criar sua réplica, estabelecendo novamente um estado de equilíbrio sobre o assunto.

## 5. Trabalhos relacionados

Os seguintes trabalhos podem ser associados a esta proposta por utilizarem os textos produzidos em Ambientes Virtuais de Ensino e Aprendizagem (AVEA) para recuperação de informações necessárias para responder outras questões.

Em [Iandoli et al. 2014] é apresentada a ferramenta *Debate Dashboard*. Essa ferramenta é integrada ao sistema *Cohere*. No artigo são apresentados argumentos e evidências empíricas que mostram que as ferramentas para ambientes colaborativos com suporte computacional para visualização de argumentações podem ser desenvolvidas para apresentar mapas conceituais sobre as interações realizadas pelo participante com o objetivo de fornecer a cada um a possibilidade de acompanhar o seu desenvolvimento dentro do debate, quais foram suas contribuições, quais relações foram realizadas com os argumentos feitos por ele. Importante ressaltar, que o *Debate Dashboard* auxilia na construção de deliberações coletivas, ou seja, fornece suporte para retirada de afirmações relevantes para o tema em debate que sejam suficientes para criar um entendimento coletivo do que foi discutido.

O trabalho de [Azevedo et al. 2011] apresenta a aplicação do software *MineraFórum*, que realiza uma análise qualitativa das mensagens enviadas por alunos em fóruns de discussão. O *MineraFórum* utiliza técnicas de mineração de textos, e tem como foco principal analisar a relevância das mensagens enviadas pelo aluno em relação ao tema em discussão.

A relação entre os trabalhos de [Iandoli et al. 2014] e [Azevedo et al. 2011] e o proposto neste artigo está na aplicação de técnicas reconhecidas para extração de informações em textos produzidos em AVEA. [Azevedo et al. 2011] analisa textos produzidos em Português e utiliza técnicas de pré-processamento de texto para reduzir o *corpus* de análise. Já [Iandoli et al. 2014] utiliza técnicas de Recuperação de Informações para tratar os textos e criar mapas conceituais que apresentam as

contribuições dos alunos no debate. Apesar de nossa proposta ser aplicada a um ambiente específico para coleta dos dados e validação, a utilização da arquitetura Multiagentes nos possibilitou modelar a camada de inteligência de forma flexível para que a mesma possa ser acoplada em qualquer AVEA, desde que os protocolos de comunicação sejam respeitados e apenas textos sejam enviados para processamento.

## 6. Considerações Finais

Este trabalho apresentou a modelagem conceitual de um Sistema MultiAgente (SMA) capaz de auxiliar o processo de mediação pedagógica em Ambientes Virtuais de Ensino e Aprendizagem, por meio da análise e recuperação de informações sobre os textos produzidos durante as interações.

A fim de validar a proposta, um protótipo funcional de parte do SMA, a ferramenta Grupos de Similaridade, composta pelo Agente de Clusterização e pelo Agente de Pré-processamento, foram implementados. O Agente de Pré-processamento é composto pelos processos responsáveis por preparar a corpus de análise que serão analisados pelo Agente de Clusterização a fim de criar os grupos solicitados. O Agente de Clusterização é composto pelo algoritmo de clusterização *K-Means* combinado com a técnica de similaridade de textos – Similaridade de Cossenos.

O protótipo desenvolvido foi aplicado ao Ambiente Virtual de Aprendizagem Debate de Teses, com o objetivo de coletar dados de interações reais realizadas em 2013. Os textos analisados foram produzidos por 21 alunos, e esses escreveram seus posicionamentos e argumentados para fundamentá-los sobre 3 teses distintas. Dessa forma, foram analisados um total de 63 textos. Os resultados encontrados com a aplicação do protótipo desenvolvido mostram-se satisfatórios e estão em consonância com as Mediações Pedagógicas identificadas e apresentadas na seção 4, pois duas das 16 mediações elicitadas são plenamente apoiadas pelo protótipo, comprovando a importância da construção de ferramentas de suporte para prática da mediação pedagógica aliada a educação à distância.

Como proposta de trabalhos futuros, pretende-se realizar a implementação completa do SMA e utilizá-lo como uma camada de inteligência integrada a dois ou mais AVEA para coleta e análise dos dados e posterior validação.

## Referências

- Afonso, S., Bick, E., Haber, R. e Santos, D. (2001). Floresta sintá(c)tica: um 'trebank' para o português. *Actas do XVII Encontro Nacional da Associação Portuguesa de Linguística (APL)*, páginas 533-545.
- Azevedo, B. F. T., Behar, P. A., e Reategui, E. B. (2011). Análise das mensagens de fóruns de discussão através de um software para mineração de textos. In *Anais do 22º Simpósio Brasileiro de Informática na Educação*, volume 1, páginas 20–29.
- Carvalho, M. J. S., Nevado, R. N., e Menezes, C. S. d. (2005). Arquiteturas pedagógicas para educação à distância: Concepções e suporte telemático. In *Anais do 16º Simpósio Brasileiro de Informática na Educação*, volume 1, páginas 351–360.
- Coppin, B. *Artificial Intelligence Illuminated*. Jones & Bartlett Learning, 2004.

- Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, 38(1):188–230.
- Feldman, R. e Sanger, J. (2006). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York, NY, USA.
- Iandoli, L., Quinto, I., De Liddo, A., e Buckingham Shum, S. (2014). Socially augmented argumentation tools: Rationale, design and evaluation of a debate dashboard. *Int. J. Hum.-Comput. Stud.*, 72(3):298–319.
- Jennings, N. R. *On agent-based software engineering*. Artificial Intelligence, v. 117, n. 2, p. 277-296, 2000.
- Manning, C. D., Raghavan, P., e Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.
- Mason, O. e Grove-Stephensen, I. (2002). Automated free text marking with paperless school. Loughborough University.
- Michels, A. B. (2014). Do fazer ao compreender no contexto da educação a distância: Uso de arquiteturas pedagógicas no processo de empreender. Dissertação de Mestrado, Programa de Pós-Graduação em Educação – UFRGS.
- Nevado, R. N., Menezes, C. S. d., e Vieira Jr, R. R. M. (2011). Debate de teses – uma arquitetura pedagógica. In *Anais do 22º Simpósio Brasileiro de Informática na Educação*, páginas 820–829.
- Panceri, S. S. e Menezes, C. S. d. (2014a). Alpes: Um sistema multiagente para análise de produções textuais no contexto de um debate de teses. In *Anais do 25º Simpósio Brasileiro de Informática na Educação*, volume 1.
- Panceri, S. S. e Menezes, C. S. d. (2014b). O suporte computacional como auxílio a mediação pedagógica em um debate de teses. In *TISE'14*, volume 1.
- Ramozzi-Chiarottino, Z. (1988). *Psicologia e Epistemologia Genética de Jean Piaget*. São Paulo: EPU.
- Reis, L. P. Coordenação em Sistemas Multi-Agente: Aplicações na Gestão Universitária e Futebol Robótico. 2003. Tese de Doutorado, Faculdade de Engenharia da Universidade do Porto. Portugal.
- Salton, G., Wong, A., e Yang, C.-S. (1975). *A vector space model for automatic indexing*. Communications of the ACM.
- Wooldridge, M. *An introduction to multiagent systems*. John Wiley & Sons, 2009.