



Universidade Federal do Rio de Janeiro



Eu-Tu

**O EMPREGO DA CLASSIFICAÇÃO AUTOMÁTICA DE
MENSAGENS EM FÓRUMS ELETRÔNICOS DE DISCUSSÕES
PARA ANÁLISE DO PROCESSO DE ENSINO E
APRENDIZAGEM CENTRADO EM INTERAÇÕES**

George Alex Fernandes Gomes

DISSERTAÇÃO DE MESTRADO



Instituto de Matemática



Instituto Tércio Pacitti de Aplicações
e Pesquisas Computacionais

Rio de Janeiro
2012

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
INSTITUTO TÉRCIO PACCITI
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

GEORGE ALEX FERNANDES GOMES

EU-TU: O EMPREGO DA CLASSIFICAÇÃO
AUTOMÁTICA DE MENSAGENS EM FÓRUNS
ELETRÔNICOS DE DISCUSSÕES PARA ANÁLISE DO
PROCESSO DE ENSINO E APRENDIZAGEM
CENTRADO EM INTERAÇÕES

Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Informática,
Instituto de Matemática e Instituto Tércio Pacciti,
Universidade Federal do Rio de Janeiro, como
requisito parcial à obtenção do título de Mestre
em Informática

Orientador: Marcos da Fonseca Elia

Rio de Janeiro
2012

G636 Gomes, George Alex Fernandes.

Eu-Tu: o emprego da classificação automática de mensagens em fóruns eletrônicos de discussões para análise do processo de ensino e aprendizagem centrado em interações/
George Alex Fernandes Gomes – Rio de Janeiro, PPGI/IM/iNCE/UFRJ, 2012.
114 fl; il.

Dissertação (Mestrado em Informática – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Instituto Tércio Pacitti de Pesquisa e Aplicações Computacionais (Núcleo de Computação e Eletrônica), 2012.

Orientador: Marcos da Fonseca Elia

1. Aprendizagem de Máquina – Teses. 2. Classificação Automática de Textos – Teses. 3. Análise de Interações – Teses. 4. Processamento da Linguagem Natural – Teses. 5. Avaliação da Aprendizagem – Teses. I. Marcos da Fonseca Elia (Orient.). II. Universidade Federal do Rio de Janeiro, Instituto de Matemática. Instituto Tércio Pacitti. III. Título.

CDD.

GEORGE ALEX FERNANDES GOMES

**EU-TU: O EMPREGO DA CLASSIFICAÇÃO
AUTOMÁTICA DE MENSAGENS EM FÓRUNS
ELETRÔNICOS DE DISCUSSÕES PARA ANÁLISE
DO PROCESSO DE ENSINO E APRENDIZAGEM
CENTRADO EM INTERAÇÕES**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Matemática e Instituto Tércio Pacciti, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Informática.

Aprovada em 28 de fevereiro de 2012.

Prof. Marcos da Fonseca Elia, PhD, PPGI/UFRJ (presidente)

Profa. Maria Luiza Machado Campos, PhD, PPGI/UFRJ

Profa. Ana Maria de Carvalho Moura, D. Ing, LNCC

Ao Senhor Deus da minha força e salvação e a todos aqueles de
quem também aprendi:
familiares, professores, colegas e alunos.

Agradecimentos

A Deus seja a glória e a honra por todo este trabalho. Foi por seu grande amor, poder e graça que consegui chegar até aqui. A cada dia, me sustentou e me deu forças para continuar, revelando sua grande fidelidade para com aqueles que nele confiam.

Agradeço a meus pais, José e Sueli, de quem recebi os primeiros ensinamentos e motivações para prosseguir estudando, por jamais terem medido esforços para que eu me tornasse um cidadão íntegro.

Agradeço a minha esposa Glória e ao meu filho Lucas pelo amor, compressão e paciência que dedicaram a mim.

Agradeço ao Professor Marcos Elia, pela orientação com esmero e de alto padrão. Não apenas por seu grande conhecimento, mas também pelo companheiro que é.

Agradeço aos meus chefes, em especial ao Major Alexandre e ao TC Castañon, que souberam entender as minhas necessidades e me apoiaram sempre.

Agradeço a Bete, Aroldo, Luiza, Paulo, Lucirene e ao "Seu Mário", que sempre se mantiveram torcendo por mim.

Agradeço aos amados irmãos em Cristo da Assembleia de Deus Filadelfia da Freguesia que me sustentaram com suas orações.

Agradeço aos amigos do CTE_x e do NCE, que acompanharam de perto esta minha jornada.

A todos esses e aos demais que prestaram apoio e incentivo, agradeço de coração e peço a Deus que sejam abundantemente abençoados a cada dia.

De tudo o que se tem ouvido, o fim é: Teme a Deus, e guarda os seus mandamentos; porque isto é o dever de todo o homem.

Eclesiastes 12:13

Resumo

GOMES, George Alex Fernandes. **Eu-Tu: o emprego da classificação automática de mensagens em fóruns eletrônicos de discussões para análise do processo de ensino e aprendizagem centrado em interações**. Rio de Janeiro, 2012 (Mestrado em Informática) - PPGI, Instituto de Matemática, Instituto Tércio Pacitti de Pesquisas e Aplicações Computacionais. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012.

Em um processo de ensino e aprendizado complexo as possibilidades de ensinar e aprender são ampliadas, uma vez que professores e alunos passam a ser vistos igualmente como fonte de conhecimento inovador. Neste contexto, as interações dos indivíduos destacam-se como o ponto de partida para a construção da nova realidade social, dando origem a uma nova forma de ensinar e aprender que chamamos de “processo de ensino e aprendizado centrado nas interações”. O presente trabalho explora uma das possíveis formas de promover esse processo em um contexto de ensino a distância, através de uma forma indireta de mediação, que se apoia na discussão escrita em um fórum eletrônico. Mais especificamente, aplicamos técnicas de processamento de linguagem natural para desenvolver um sistema automático, que denominamos de *Eu-Tu*, capaz de analisar oportunamente as interações e produzir indicadores de avaliação a partir das mensagens produzidas.

Palavras-chave: Classificação Automática de Textos. Análise de Interações. Aprendizagem de Máquina. Processamento da Linguagem Natural.

Abstract

GOMES, George Alex Fernandes. **Eu-Tu: o emprego da classificação automática de mensagens em fóruns eletrônicos de discussões para análise do processo de ensino e aprendizagem centrado em interações**. Rio de Janeiro, 2012 (Mestrado em Informática) - PPGI, Instituto de Matemática, Instituto Tércio Pacitti de Pesquisas e Aplicações Computacionais. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012.

In a complex process of teaching and learning, possibilities of teaching and learning are enhanced, since teachers and students are also seen as a source of innovative knowledge. In this context, the interactions of individuals stand out as the starting point for the construction of new social reality, giving rise to a new way of teaching and learning we call "interactions-centered teaching and learning process." This work explores one of the ways to improve this process in a context of distance education, through an indirect form of mediation, which is based on the discussion written in an electronic forum for discussions. More specifically, we apply techniques from natural language processing to develop an automated system, we call Eu-Tu, able to analyze the interactions and produce timely assessment indicators from the messages produced.

Keywords: Automatic Text Classification. Interaction Analysis. Machine Learning. Natural Language Processing.

Lista de Figuras

Figura 1. Referencial teórico	30
Figura 2. O universo geosocial da Internet.	32
Figura 3. Tipos de conversação pela Internet.	34
Figura 4. Evolução do número de pessoas com acesso à Internet no Brasil.	35
Figura 5. A evolução da comunicação..	44
Figura 6. Classificação supervisionada.	51
Figura 7. Diagrama representativo do algoritmo Árvore de Decisão.....	52
Figura 8. Avaliação Mediadora.	56
Figura 9. Modelo básico do Eu-Tu.....	56
Figura 10. Eu-Tu instanciado como um Sistema de Perguntas e Respostas.	58
Figura 11. Diagrama do Sistema de Perguntas e Respostas.....	59
Figura 12. Eu-Tu instanciado como um Sistema de Perguntas e Respostas, com Combinação Social.	60
Figura 13. Distribuição dos Estilos de Aprendizagem dos alunos do PGTIAE	64
Figura 14. Uma instância do modelo básico do <i>Eu-Tu</i> aplicada em Fóruns Eletrônicos de Discussões.	69
Figura 15. Código para medir a precisão dos classificadores.....	73
Figura 16. Fórmula da precisão para uma classe "A" hipotética.	74
Figura 17. Código para criação de um classificador Bayesiano binário	74
Figura 18. Código para criação dos conjuntos de treinamento e de teste.....	75
Figura 19. Código para extrair as palavras mais significativas	76
Figura 20. Influência do tamanho do conjunto de treinamento no desempenho do classificador Bayesiano binário - EXPERIMENTO III.....	80
Figura 21. Distribuição das categorias - EXPERIMENTO IV	82
Figura 22. Influência do número de categorias na precisão do classificador Bayesiano multiclasse - EXPERIMENTO IV	83
Figura 23. Particionamento usado para criar o <i>corpus</i> do classificador da categoria "Saudação".....	84
Figura 24. Código para construção do conjunto de classificadores binários.....	85
Figura 25. Código para determinar o voto dos classificadores.....	85
Figura 26. Diagrama do esquema de votação para o conjunto de classificadores binários	86
Figura 27. Código para determinar a precisão individual dos classificadores binários .	86

Lista de Quadros

Quadro 1. Desempenho dos classificadores - EXPERIMENTO I	75
Quadro 2. Desempenho dos classificadores - EXPERIMENTO II	77
Quadro 3. Variação do tamanho do conjunto de treinamento - EXPERIMENTO III....	79
Quadro 4. Distribuição das mensagens por categorias - EXPERIMENTO IV	81
Quadro 5. Fases de aprendizagem, segundo Gunawardena, Lowe e Anderson (1997).	110
Quadro 6 - Categorias propostas por Lopes (2007).....	111
Quadro 7- Relação entre as categorias de Lopes (2007) e as fases de Gunawardena, Lowe e Anderson.....	113

Lista de Abreviaturas e Siglas

AHS	Sistema Hipermídia Adaptativo (do inglês, <i>Adaptive Hypermedia System</i>)
ANATEL	Agência Nacional de Telecomunicações
AVA	Ambiente Virtual de Aprendizagem
CSCL	Aprendizagem Colaborativa (do inglês, <i>Computer Supported Collaborative Learning</i>)
CSIL	Aprendizagem Investigativa Apoiada pelo Computador (do inglês, <i>Computer Supported Inquiry Learning</i>)
EAD	Educação a Distância
EB	Exabytes
GB	Gigabytes
GINAPE	Grupo de Informática Aplicada à Educação
ILS	Index of Learning Styles
IP	Internet Protocol
k-NN	K-Nearest Neighbor
LSA	Análise Semântica Latente (do inglês, <i>Latent Semantic Analysis</i>)
LSI	Learning Style Inventory
MBTI-1	Myers-Briggs Type Indicator 1
MINTS	Minnesota Internet Traffic Studies
NCE	Núcleo de Computação Eletrônica
NLTK	Natural Language Toolkit
Pii	Plataforma Interativa para Internet
SCM	Serviço de Comunicação Multimídia
SVM	Máquinas de Suporte Vetorial (do inglês, <i>Support Vector Machines</i>)
TIC	Tecnologias da Informação e da Comunicação
UFRJ	Universidade Federal do Rio de Janeiro
UNICAMP	Universidade de Campinas
ZB	Zettabyte

SUMÁRIO

CAPÍTULO 1: INTRODUÇÃO	15
1.1 CONTEXTO E MOTIVAÇÃO	15
1.1.1 O PROCESSO DE ENSINO E APRENDIZAGEM CENTRADO EM INTERAÇÕES	16
1.2 CARACTERIZAÇÃO DO PROBLEMA	18
1.3 HIPÓTESE.....	19
1.4 OBJETIVOS	19
1.4.1 OBJETIVO GERAL.....	19
1.4.2 OBJETIVO ESPECÍFICO.....	22
1.5 PREMISSA	23
1.6 TRABALHOS RELACIONADOS	23
1.7 ORGANIZAÇÃO DA PESQUISA	28
CAPÍTULO 2: REFERENCIAL TEÓRICO.....	30
2.1 OS CAMINHOS PARA O EMBASAMENTO TEÓRICO	30
2.2 A REALIDADE QUE VIVEMOS	31
2.3 DADOS, DADOS E MAIS DADOS.....	31
2.4 MAIS IMPORTANTE QUE INFORMAR, É TRANSFORMAR.....	36
2.5 NOVAS CARACTERÍSTICAS DOS APRENDIZES.....	38
2.6 COMO TIRAR PROVEITO DAS NOVAS CARACTERÍSTICAS DOS APRENDIZES	40
2.6.1 USANDO INFORMAÇÕES.....	40
2.6.2 USANDO INTERAÇÕES.....	41
2.7 O PONTO DE PARTIDA.....	42
2.7.1 OBSTÁCULOS À INOVAÇÃO.....	42
2.7.2 DISCUTIR EM GRUPO PODE LEVAR À INOVAÇÃO	43
2.7.3 PREFERÊNCIA PELO TEXTO	44

2.8	EXPLORANDO AS INTERAÇÕES ESCRITAS: AS RIQUEZAS LATENTES DAS MENSAGENS	45
2.8.1	FÓRUMS ELETRÔNICOS DE DISCUSSÕES: AMBIENTES CAPAZES DE PROMOVER O PENSAMENTO CRÍTICO	46
2.8.2	FÓRUMS ELETRÔNICOS DE DISCUSSÕES: AMBIENTES PROPÍCIOS PARA AVALIAÇÃO	47
2.8.3	É PRECISO MODERAÇÃO.....	48
2.9	ANÁLISE DE CONTEÚDO	48
2.9.1	ANÁLISES QUANTITATIVA E QUALITATIVA.....	49
2.9.2	CLASSIFICAÇÃO AUTOMÁTICA DE DOCUMENTOS.....	50
2.9.3	ALGORITMOS DE CLASSIFICAÇÃO	51
2.9.3.1	O ALGORITMO ÁRVORE DE DECISÃO	51
2.9.3.2	O ALGORITMO DE NAÏVE BAYES	53
2.9.3.3	O ALGORITMO DE MÁXIMA ENTROPIA	53
2.9.4	NECESSIDADE DE AUTOMATIZAÇÃO	54
	CAPÍTULO 3: CONTEXTUALIZAÇÃO CIRCUNSTANCIADA.....	55
3.1.	A EVOLUÇÃO DA PESQUISA.....	55
3.1.1.	O MODELO BÁSICO DO EU-TU.....	55
3.1.2.	O EU-TU COMO UM SISTEMA COLABORATIVO DE PERGUNTAS E RESPOSTAS.....	57
3.1.3.	ESTUDOS COM ESTILOS DE APRENDIZAGEM	61
3.1.4.	MUDANÇA DE RUMO	66
	CAPÍTULO 4: A IMPLEMENTAÇÃO DO MÓDULO DE AVALIAÇÃO	68
4.1.	PROMOVENDO O PROCESSO DE ENSINO E APRENDIZADO CENTRADO EM INTERAÇÕES ATRAVÉS DA TROCA DE MENSAGENS EM FÓRUMS ELETRÔNICOS DE DISCUSSÕES	68
4.2.	O MODELO DE AVALIAÇÃO ESCOLHIDO PARA O EU-TU	70
4.3.	DIFERENCIAL E OPORTUNIDADES DE MELHORIA.....	71
4.4.	DESENVOLVIMENTO DO MÓDULO DE AVALIAÇÃO	72

4.4.1.	ETAPA I: DETERMINAÇÃO DO ALGORITMO DE CLASSIFICAÇÃO	72
4.4.1.1.	EXPERIMENTO 1: ANÁLISE DE DESEMPENHO.....	73
4.4.1.2.	EXPERIMENTO 2: INFLUÊNCIA DO ESPAÇO DE ATRIBUTOS	76
4.4.1.3.	A ESCOLHA DO CLASSIFICADOR	77
4.4.2.	ETAPA II: CONSTRUÇÃO DO CLASSIFICADOR	78
4.4.2.1.	EXPERIMENTO III: INFLUÊNCIA DO TAMANHO DO CONJUNTO DE TREINAMENTO	78
4.4.2.2.	EXPERIMENTO IV: INFLUÊNCIA DO NÚMERO DE CATEGORIAS	80
CAPÍTULO 5: RESULTADOS E ANÁLISE		84
CAPÍTULO 6: CONCLUSÕES E TRABALHOS FUTUROS		89
6.1.	VISÃO GERAL	89
6.2.	PROBLEMAS ENCONTRADOS.....	90
6.3.	CONTRIBUIÇÕES	91
6.4.	TRABALHOS FUTUROS	93
6.5.	CONSIDERAÇÕES FINAIS	94
REFERÊNCIAS		Erro! Indicador não definido.
ANEXO 1.....		110
ANEXO 2.....		111
ANEXO 3.....		113

CAPÍTULO 1: INTRODUÇÃO

1.1 CONTEXTO E MOTIVAÇÃO

A teoria da complexidade considera que conhecimento não é somente aquilo que já se sabe, mas também o que ainda não sabemos e que fornece um leque de opções de respostas. Com isso, as interações dos indivíduos passam a ser o novo ponto de partida para a construção da realidade social (ELIA, 2008; OLIVEIRA, 1997). Uma evidência desse princípio na sociedade é o rápido crescimento do fluxo de informações, ocorridos nos últimos cinquenta anos, provocado pelo fenômeno da globalização.

Do ponto de vista da educação, a grande mudança provocada pela complexidade foi tornar professores e alunos igualmente fontes de conhecimento inovador. Esse efeito pode ser percebido pelo crescente emprego das Tecnologias da Informação Aplicada à Educação — TIAE, para o desenvolvimento acelerado de redes de ensino e aprendizado colaborativos e distribuídos, no tempo e no espaço. Com isso, acrescenta-se uma nova estigmergia à linha pedagógica do processo de ensino e aprendizagem (SIEMENS, 2005; HUMMEL et al., 2006; TATTERSALL et al, 2004), o que força os papéis desempenhados por professores e alunos a não mais se limitarem aos tradicionais modelos educacionais, onde geralmente atuavam de forma estática e permanente (BONABEAU, 1999; TATTERSALL et al, 2004).

Os alunos não mais têm seus professores como a única fonte de conhecimento (modelo *broadcasting*¹ – 1:N) e almejam impor seus próprios ritmos de aprendizagem. Não querem apenas "absorver" o que lhes é apresentado, mas sim "explorar" o conhecimento (modelo N:N) (HAY, 2000). Desejam oportunidades de aprendizagem auto-direcionadas, realizadas em ambientes interativos e que lhes proporcionem um ensino mais proativo, baseado em práticas educacionais compatíveis com as tecnologias presentes no cotidiano da sociedade atual e de acordo com suas preferências individuais (GLENN, 2000; GELLER, TAROUCO, FRANCO, 2004; KINSHUK, GRAF, 2007). Ao professor, cabe tirar proveito dessas mudanças e dar intencionalidade às atividades educacionais, tornando-se assim, um projetista do saber, capaz de indicar o conteúdo mais apropriado e apresentá-lo por meio de

¹ Modelo de comunicação em que um transmissor emite mensagens para diferentes receptores. Também chamado de *multicast*.

uma linguagem adequada para viabilizar o aprendizado de seus alunos (ELIA, 2008).

O professor precisa considerar que o conhecimento não pode ser transferido de uma mente para outra, como se fosse algo concreto. As ideias devem ser pensadas na mente dos alunos e as experiências, revividas por eles. Isto lhes trará motivação para buscar e compartilhar novos conhecimentos. Nesse sentido, Gregory (1993), chega a afirmar que "ensina melhor quem menos ensina". Portanto, ao invés de tentar transferir conhecimento para seus alunos, o professor deve estimular e orientar a busca do conhecimento, apresentando mais estratégias do que técnicas (POZO, 1998). Assim, os alunos serão forçados a aprenderem de forma autônoma e a participarem ativamente do processo de ensino-aprendizado (FREIRE, 2011). Naturalmente, novos papéis estão também destinados aos demais protagonistas dessa nova escola, como os gestores, pais e responsáveis, governo e Estado.

1.1.1 O PROCESSO DE ENSINO E APRENDIZAGEM CENTRADO EM INTERAÇÕES

Nos modelos de ensino-aprendizagem centrados no professor, o processo de ensino e aprendizagem é basicamente desenvolvido por dois papéis disjuntos, instrutor e instruendo, que possuem atribuições exclusivas e regras rígidas para ações: ao instrutor cabe ensinar e ao instruendo, ser instruído. Indiscutivelmente, assemelha-se ao que ocorre em uma comunicação *broadcasting*. Portanto, o professor "transmite" seus conhecimentos, para os alunos os "receber". A rigidez dessa regra é tão forte que constitui um processo de uma única via. De um lado, o "Eu-professor" envia ensinamentos para o "Tu-aluno" no cumprimento de seu dever de ensinar. Do outro, o "Eu-aluno" recebe os ensinamentos do "Tu-professor" como o esforço suficiente que o levará a aprender.

As tentativas para mudar esse cenário iniciam-se com os modelos centrados no aluno. Os mesmos papéis de instrutor e instruendo ainda são mantidos para o professor e para o aluno respectivamente. Entretanto, há uma inversão do ponto de partida das ações. É o aluno que ocupa o foco, tornando-se o referencial que delinea todo o processo. Em alguns modelos mais sofisticados, a aprendizagem em grupo é estimulada ao longo do processo, permitindo que não mais ocorra apenas em uma única via.

Ao dar uma nova definição para "conhecimento inovador", a teoria complexidade desencadeia uma série de mudanças que rompem de vez a rigidez dos papéis desempenhados nos modelos *broadcasting* e proporcionam maior flexibilidade para as interações ocorrerem.

A principal mudança ocorre nos significados de "ensinar" e "aprender". Não mais estão associados à transferência de conhecimento de uma mente para outra, como se fosse concreto, um objeto pronto e acabado, e sim às ideias que são pensadas a respeito e às experiências que são revividas.

Uma vez que todos se tornam igualmente fontes de conhecimento inovador, "aprender" traz consigo um pouco de "ensinar" e vice-versa. O instruendo ganha a oportunidade de contribuir para o ensino, enquanto o instrutor, a de poder aprender.

Naturalmente, a colaboração se torna uma atribuição desses dois papéis e, com isso, o contraste entre eles é atenuado a ponto das diferenças ficarem disfarçadas. A distinção se restringe às ações para dar intencionalidade à aprendizagem, que permanecem exclusivas ao instrutor. Assim, regras antigas são esquecidas e o que passa a valer é: ao aprender, pode-se ensinar; ao ensinar, pode-se aprender.

O instruendo precisa, então, ser mais ativo, colaborando com o aprendizado de outros. O instrutor, do alto de sua competência e sabedoria, deve, em prol da aprendizagem, considerar e tirar proveito de que aquilo que ainda não sabemos é o que potencialmente pode ser transformado em conhecimento inovador. Portanto, professores e alunos, outrora limitados a exercerem papéis com normas fixas e atribuições exclusivas, ganham liberdade para atuarem no desenrolar de suas interações, proporcionando mais dinamismo ao processo de ensino e aprendizagem.

Pelo exposto, vê-se que esse novo paradigma tem ampliado as possibilidades de ensinar e aprender, demandando a criação de um processo que considera não apenas o conhecimento bibliotecário, aquele que se desenvolve pelos tipos e quantidades de informações já consolidadas (conteúdo), mas também o conhecimento contextualizado, que é fruto das ações que se estabelecem entre os fatos e os sujeitos envolvidos (prática) (SIEMENS, 2005). Com isso, a ordem tradicional teoria-prática é invertida para prática contextualizada-teoria, por meio do eventual conhecimento necessário para criar as competências e habilidades requisitadas (ROQUE et al, 2004).

A esse novo processo colaborativo e adaptativo, em que tanto o professor quanto o aluno deixam de ser o foco das ações, e que servirá de base para o objeto desta pesquisa, denominamos de *processo de ensino e aprendizagem centrado em interações*.

1.2 CARACTERIZAÇÃO DO PROBLEMA

Com todas as mudanças trazidas pela teoria da complexidade, não há mais razões para se considerar um "Eu" separado de um "Tu", e sim de um "*Eu-Tu*" que ensina e aprende ao mesmo tempo. Daí surge a seguinte dúvida:

Como a escola e o professor podem tirar proveito das novas oportunidades trazidas pela complexidade para aprimorar o ensino e a aprendizagem em um cenário complexo, onde tudo (objetos de aprendizagem, materiais didáticos, etc..) e todos (professores, alunos, pais, comunidade vicinal, sociedade em rede, etc..) são igualmente fontes de conhecimento inovador e as interações dos indivíduos são o ponto de partida para a construção da realidade social?

Responder essa pergunta requer cuidados, pois diferentes aspectos devem ser considerados, podendo desencadear uma infinidade de abordagens. Por exemplo, um primeiro passo, é considerar que as interações devem ser efetivas. Assim, o questionamento inicial seria abordado quanto aos aspectos relacionados à natureza das interações:

- Interação induzida pelo professor através de uma atividade didática em grupo (tarefa, projeto, etc.);
- Interação colaborativa que gera uma dependência entre os aprendizes no sentido de que o sucesso de uns dependa do sucesso dos outros;
- Interação motivacional e contextualizada que se desenvolve a partir de demandas reais e relacionadas com o dia a dia dos aprendizes.

Enfim, esses seriam apenas alguns exemplos. Como já foi dito, cada um desses aspectos, tomados de forma isolada ou combinada levariam a uma variedade de abordagens, estratégias e modelos pedagógicos.

Podemos, então, enunciar que o problema deste trabalho de pesquisa está relacionado aos seguintes questionamentos:

- *Como determinar as variáveis que exercem influência sobre as interações de alunos e professores?*
- *Como distinguir as variáveis que favorecem das que não favorecem a ocorrência de interações desejadas?*
- *Como fazer uso dessas variáveis para produzir boas oportunidades de ensino e de aprendizagem?*

1.3 HIPÓTESE

Com base nos questionamentos levantados, formulamos a seguinte hipótese para este trabalho de pesquisa:

É possível construir uma instância de um modelo computacional para promover o processo de ensino e aprendizagem centrado em interações.

1.4 OBJETIVOS

A mente humana funciona ativada por motivações. Desmotivado, o ser humano não ouve nem vê nada. Portanto, quando professores e alunos se encontram adequadamente motivados, agem com efetividade.

Todavia, para que não produzam efeitos desastrosos, que tornam alunos insatisfeitos e professores frustrados, as motivações devem resultar ensino e aprendizagem contextualizados.

1.4.1 OBJETIVO GERAL

Uma variável legítima do processo de ensino e aprendizagem centrado nas interações é a colaboração. Ele considera que os alunos não aprenderão tudo com seus professores, mas, se colaborarem, poderão aprender uns com os outros. Não admite que o conhecimento seja algo transferido de um instrutor para um instruendo, que absorve passivamente o conteúdo que lhe é apresentado, e sim construído coletivamente, fruto de diferentes interações sociais, como negociação, cooperação, deliberação e argumentação. Portanto, pode-se definir que uma

interação é efetiva para o ensino e a aprendizagem quando motiva e apoia a colaboração, na medida em que as ideias sejam pensadas nas mentes das pessoas e as experiências revividas por cada uma delas.

Mas, o que fazer para as interações serem efetivas? Em primeiro lugar, é preciso que sejam fornecidos subsídios para aprender a colaborar, até que, posteriormente, ocorra colaboração para aprender (GOMES NETO, GOMES, TEDESCO, 2003). Os alunos devem ser motivados a colaborar não por obrigação, mas por desejo próprio. Uma das melhores maneiras de despertar esse desejo é torná-los cientes de suas limitações e das oportunidades que os cercam. Quando percebem que podem se beneficiar com o que está sendo ensinado, seus níveis de motivação para aprender tornam-se bem mais elevados (GREGORY, 1993). Entretanto, precauções são necessárias. Não se deve promover uma motivação do "Eu" sem o "Tu". Os alunos precisam deixar para trás a visão egoísta de buscar novos conhecimentos apenas para si. É necessário que estejam conscientes de que seus conhecimentos, e dúvidas, podem servir para a aprendizagem de outros.

A motivação também não pode ser "mercenária", aquela que faz tudo o que os alunos gostam e desejam. Nem sempre o que eles querem é o que realmente precisam. Devem ser provocados a sair de suas zonas de conforto, para desenvolverem novas habilidades. É tarefa do professor, levar seus alunos a pensar por si mesmos, a ser disciplinados e a agir por deliberação própria. O professor deve motivar e dirigir os atos da aprendizagem e, de modo geral, não deve dizer nem fazer para o aluno nada que ele possa fazer por si mesmo, a não ser por algumas exceções, em que houver maiores dificuldades ou que necessitar de incentivo. Mesmo assim, essa ajuda deverá ser mantida apenas até o aluno ter condições de prosseguir por conta própria. Caso contrário, correrá o risco de formar "deficientes" intelectuais, que não conseguem aproveitar as oportunidades de ensino e de aprendizagem. Será bem melhor que se empenhem mais a questionar respostas do que a responder perguntas (HENDRICKS, 1991).

Outra forma de motivar os alunos é estruturar corretamente a experiência de ensino e de aprendizagem, dando intencionalidade às ações que serão realizadas (HENDRICKS, 1991). Se não houver intencionalidade, possivelmente os alunos seguirão por caminhos que não os levarão a lugar algum. Ainda assim, o professor poderá usar o insucesso como alicerce para ensinar. Mas para isso, precisa ser capaz de mostrar aos alunos que as falhas cometidas e os resultados negativos são oportunidades para aprender.

É interessante que os alunos, sem exceção, poderão estar motivados a aprender, mas, provavelmente, não todos ao mesmo tempo nem da mesma forma, com o mesmo professor.

Individualmente, apresentam diferentes níveis de competência e possuem ritmos de aprendizagem e experiências de vida distintas e, com isso, não aprendem todos em igual medida, qualitativa e quantitativamente.

Há, portanto, momentos e formas específicas, para os quais, cada aluno se apresenta predisposto a aprender. É atribuição do professor descobrir esses momentos e essas formas, para ter uma percepção clara do andamento do processo de ensino e aprendizagem. Deve, portanto, conhecer bem seus alunos e deixar que o conheçam. Quando conseguir identificar e aproveitar uma dessas oportunidades e utilizar uma dessas formas, aperfeiçoará o processo de ensino e aprendizagem. Por exemplo, ao adaptar modelos e métodos tradicionais de diferentes níveis de ensino, a EAD se tornou uma ferramenta importante para a formação continuada.

Nesse aspecto, mediar e avaliar constituem tarefas fundamentais para o professor desempenhar bem seu papel. A mediação e a avaliação representam componentes das ações e reações que ocorrem no processo de ensino e aprendizagem e, quando associadas corretamente, formam um binário capaz de alavancar esse processo. Permitem que sejam levantadas e incorporadas informações para realizar regulações constantes, que levam à melhoria da qualidade do ensino, tanto individual quanto coletivo (KRAEMER, 2005).

Por sua vez, considerando um ponto de vista construtivista e interativista, ambientes que estimulam colaboração para produção de conhecimentos e, ao mesmo tempo, oferecem possibilidades de customização do ensino, viabilizam processos autônomos de aprendizagem. Mas, para que, de fato, ocorra educação ativa, é necessário que, em meio a confrontos de interesses individuais e coletivos, as direções tomadas sejam as corretas.

As ações devem ser propulsionadas e respaldadas por avaliações criteriosas e contínuas do processo de ensino-aprendizagem, capazes de dialogar com a complexidade do real, que abrange uma multiplicidade de conhecimentos, papéis e valores.

Pensando assim, este trabalho dá ênfase à avaliação como um processo essencial para promover o processo de ensino e aprendizagem centrado em interações. O seu objetivo geral é formular um modelo de ferramenta computacional, que denominamos de Eu-Tu, capaz de ser acoplada aos ambientes de EAD, para auxiliar a avaliação de interações que ocorrem em atividades colaborativas, com a expectativa de que seu emprego sirva para dar suporte às diferenças entre seus usuários e tornar a ação pedagógica mais efetiva.

1.4.2 OBJETIVO ESPECÍFICO

Tratando-se de ambientes apoiados por ferramentas computacionais que viabilizam a interatividade no processo de aprendizagem à distância, existem inúmeros recursos e resultados que podem ser considerados como indicadores de qualidade e que afetam tanto as ações dos alunos quanto as dos professores. Surgem, então, outros questionamentos:

- *Como diagnosticar se houve aprendizagem e em que fase do processo aconteceu?*
- *Como identificar a contribuição de cada participante para a execução das tarefas?*
- *Como verificar se a interação realmente contribuiu para a aquisição de conhecimento?*

Fóruns categorizados semanticamente têm servido como ferramenta alternativa para solucionar questões desse tipo e, por isso, são cada vez mais empregados no processo de avaliação em EAD/TIC (LOPES, 2007). As mensagens geradas através desses fóruns podem ser subdividas em unidades representativas e classificadas de acordo com o significado de seus conteúdos para analisar as interações que ocorreram durante as discussões (HENRI, 1992; GUNAWARDENA, 1997, LUI et al, 2004; SWIGGER, BRAZILE, 2010).

A análise das interações escritas mostra-se como uma promissora ferramenta para aprimorar o processo de ensino e aprendizagem (SWIGGER, BRAZILE, 2010), mas também *"pode ser uma tarefa árdua ou até mesmo impossível para um professor, dependendo da quantidade de interações, alunos ou sessões."* (LOPES, 2007). Além do mais, as variáveis que mais influenciam o complexo processo de ensino e aprendizagem estão bem entrelaçadas e, por isso, são difíceis de serem isoladas (VAN DER POL, 2002). Muitas informações redundantes ou desnecessárias precisam ser eliminadas para se alcançar bons resultados.

O emprego de técnicas de mineração de dados, de análise eletrônica de conteúdo e de aprendizagem de máquina mostra-se essencial para superar tais dificuldades. Por isso, decidimos aplicar técnicas de Processamento de Linguagem Natural – PLN² para avaliar automaticamente o processo de construção do conhecimento colaborativo em ambientes de EAD.

Em síntese, este trabalho explora uma das possíveis formas de promover o processo de ensino e aprendizado centrado em interações. Trata-se de uma forma indireta de avaliação,

² Em inglês, *Natural Language Processing* — NLP.

que se apoia na análise automática de mensagens trocadas em um fórum eletrônico de discussões, para identificar oportunamente os níveis de aprendizagem a partir das interações que ocorrerem.

Inicialmente, também pretendíamos verificar se a interação escrita pode estimular o desenvolvimento de novas habilidades e contribuir para a mediação do processo de ensino e aprendizagem através da recomendação de parceiros para novas interações. Entretanto, essa verificação ficou para trabalhos futuros. Neste trabalho, priorizamos a aquisição de indicadores que revelam os efeitos dos processos sociais e cognitivos durante a construção de conhecimento. Julgamos que, após termos concluído essa etapa, teremos mais subsídios para prosseguir em busca do objetivo geral desta pesquisa. Portanto:

O objetivo específico deste trabalho é desenvolver um mecanismo computacional que automatize a classificação de mensagens em fóruns eletrônicos de discussões e que seja capaz de apresentar indicadores de aprendizagem para proporcionar interações efetivas em cursos EAD. Tal mecanismo poderá servir também como um instrumento alternativo de avaliação formativa durante o processo de aprendizagem.

1.5 PREMISSA

O presente trabalho tem como premissa os resultados alcançados por Lopes (2007), que apontam para a possibilidade de se realizar a avaliação quantitativa e qualitativa relativa à aquisição de conhecimento de um grupo de pessoas a partir da classificação e do mapeamento apropriados dos significados produzidos pelas interações registradas em uma ferramenta de comunicação assíncrona usada como apoio para o desenvolvimento de uma atividade colaborativa.

1.6 TRABALHOS RELACIONADOS

Capretz (2002, 2006), Layman, Cornwell, e Williams (2006) apresentaram uma série de práticas para o ensino de Engenharia de Software. O intuito desses trabalhos foi desenvolver estratégias para entender como os alunos abordam suas tarefas de aprendizagem e, a partir daí, ajustar o ensino.

Em estudos mais recentes, Santos (2007) analisou a influência dos estilos de

aprendizagem no perfil do aluno de EAD e Adán Coello et al. (2008) utilizou estilos de aprendizagem, juntamente com a teoria do conflito sócio-cognitivo, para a construção de uma ferramenta que auxilia o professor a formar grupos para o estudo colaborativo de desenvolvimento de software. Essa ferramenta agrupa alunos com os mesmos estilos, ou complementares, na tentativa de evitar conflitos indesejáveis, que podem ocorrer devido à existência de estilos de aprendizagem incompatíveis dentro do grupo.

Hiroaki et al (2008a) também estudaram o problema de formação de grupos para o ensino de Engenharia de Software, considerando as diferenças de habilidades dos alunos. Como solução, eles desenvolveram inicialmente um sistema, chamado EtUDE, para apoiar a realização, em grupo, de exercícios de Engenharia de Software, capaz de detectar alunos que não se beneficiam com esse modo de trabalho e que precisam de instrução individualizada. Posteriormente, Hiroaki et al (2008b) propuseram o EtUDE/GO, uma evolução do EtUDE, que gera automaticamente a melhor composição dos grupos.

Alfonseca et al. (2006) conduziram um estudo que mostrou como os benefícios de se considerar os estilos de aprendizagem para a criação de um modelo de usuário (*User Model*) podem ser ampliados em ambientes de aprendizagem colaborativa, como um ponto-chave na formação de grupos mais produtivos. Com esse intuito, as informações a respeito das interações dos alunos foram utilizadas para a extração automática de regras de agrupamento.

Paredes e Rodriguez (2006) utilizaram o modelo de estilos de aprendizagem de Felder-Silverman e seu questionário, o *Index of Learning Styles* (ILS), com a finalidade de classificar estudantes para tarefas colaborativas. Eles observaram que algumas dimensões do modelo afetaram a qualidade das tarefas realizadas. Dessas observações, incorporaram ao sistema TANGOW/WOTAN, novas regras para a formação de grupos.

Kinshuk e Graf (2007) apontaram que a modelagem dos alunos é um desafio para o desenvolvimento de ambientes virtuais de aprendizagem adaptativa. Popescu e Badica (2009) propuseram um sistema de aprendizagem adaptativo, chamado WELSA (*Webbased Educational system with Learning Style Adaptation*), baseado não apenas em um único modelo de estilos de aprendizagem, mas em um conjunto de características extraídas de vários modelos apresentados na literatura, que denominaram de Modelo Unificado de Estilo de Aprendizagem (*Unified Learning Style Model* - ULSM). O objetivo pedagógico desse sistema é recomendar aos estudantes objetos e rotas de aprendizagem mais adequados aos seus estilos. Com essa finalidade, são utilizadas técnicas de classificação e anotação adaptativa para inserir, eliminar, distribuir e mover os ponteiros (*links*) para objetos de aprendizagem

específicos.

O sucesso dos sistemas de recomendação na área de comércio eletrônico tem inspirado algumas pesquisas para a adaptação em ambientes eletrônicos de aprendizagem. Al Hamad, Yacob e Al-Zoubi (2008), por exemplo, integraram as informações a respeito dos estilos de aprendizagem a um sistema de recomendação personalizado (*User Modeled Personalized E-learning Recommender System* - UMPERS), para auxiliar professores no emprego de metodologias mais adequadas aos estilos de seus alunos. Esse sistema propõe um conjunto de questões e aplica técnicas de análise semântica latente (*Latent Semantic Analysis* - LSA) às respostas dos alunos, para identificar seus níveis de conhecimento e estilos de aprendizagem. Em seguida, através de um conjunto de regras, estabelecidas por um subsistema especialista, uma lista de materiais é recomendada a cada aluno.

Candotti et al. (2006) propuseram a implementação de um Sistema Hipermídia Adaptativo (*Adaptive Hypermedia System* - AHS) para EAD, via Web, baseado no estilo de aprendizagem dos alunos. Através de um teste elaborado por Bariani (1998), o sistema identifica os estilos de aprendizagem e exibe a forma de explanação dos conteúdos e o tipo de atividade mais adequada a ser apresentada.

Para o processo de ensino e aprendizagem centrado nas interações, onde todos somos potencialmente considerados como fontes de conhecimento inovador em uma realidade complexa, a recomendação de materiais específicos ou rotas de aprendizagem não parece ser a mais propícia. Ademais, os resultados das pesquisas relacionadas confirmam que é um equívoco acreditar que seja possível desenvolver uma técnica única de ensinar, capaz de abarcar todos os alunos ao mesmo tempo. Cada indivíduo recebe e processa informações de maneira peculiar. Para que haja uma instrução efetiva, é preciso que as técnicas de ensino sejam adequadas aos diferentes estilos de aprender dos alunos. Dentre outras coisas, isso significa incorporar atividades que estimulem a reflexão e a discussão, criar oportunidades para desenvolver a percepção individual, desafiar os alunos com problemas práticos e abstratos e encorajá-los a enxergar tanto de forma específica quanto abrangente.

Pretendemos também que o *Eu-Tu* seja uma ferramenta para apoiar atividades didáticas colaborativas. Entretanto, existem pontos distintos entre a proposta deste trabalho para os demais relacionados. Apesar de considerarmos que a recomendação de alunos, para interagirem como parceiros de aprendizagem, seja uma forma efetiva de mediação, o foco de nossa estratégia inicial não está na formação de grupos, e sim na avaliação do processo de ensino e aprendizagem por meio da análise de interações escritas. Portanto, nossa proposta

está um passo a frente. Ao invés de nos preocuparmos como se dará a mediação propriamente dita, atentamos para os efeitos provocados por ela. Nosso objetivo é a criação de um mecanismo que produza indicadores que permitam avaliar o resultado das interações ocorridas. Processos autônomos de aprendizagem, capazes de potencializar novas interações, poderão ser viabilizados por esses indicadores. Inclusive, um exemplo a ser verificado é o emprego de regras para a recomendação de parceiros gerada a partir dos indicadores produzidos.

A Aprendizagem Colaborativa Mediada por Computador (do inglês, *Computer Supported Collaborative Learning - CSCL*), tem se tornado uma prática educacional cada vez mais empregada. Os construtivistas cognitivos afirmam que o emprego de ambientes de CSCL promovem o aprendizado devido à explicitação dos elementos de conhecimento individuais (recuperados da memória) e a consequente reorganização desses elementos ao longo das interações sociais. Os sócios-construtivistas argumentam que os ambientes CSCL promovem um processo de colaboração em que os significados são negociados e os conhecimentos co-construídos coletivamente (DE WEVER, SCHELLENS, VALCKE, 2006).

A despeito do ponto de vista a considerar, ambos destacam a importância da interação para a aprendizagem colaborativa. O estudo dessas interações, confinadas nas transcrições das discussões de fóruns eletrônicos, tem sido objeto de muitas pesquisas. Os fóruns eletrônicos constituem repositórios de informações preciosas e inexploradas. Do ponto de vista educacional, essas informações podem revelar o processo de ensino e aprendizagem e as características dos alunos e dos professores. Portanto, a análise de conteúdo constitui um instrumento eficaz para aprimorar o processo de ensino e aprendizagem.

Uma etapa da análise de conteúdo compreende a classificação de mensagens de acordo com um tipo específico de informação relacionada às interações que ocorrem durante as discussões (ROURKE et al, 2001), como categorias que expressam aspectos de afetividade, cognição e coesão (ROURKE et al, 1999). Especialmente em ambientes de EAD, onde a qualidade da participação dos alunos é considerada de fundamental importância, a análise de conteúdo pode contribuir para avaliação da aprendizagem. É possível eliminar mensagens propositalmente enviadas para criar uma falsa impressão de participação. Nesse caso, os textos são classificados de acordo com uma medida de qualidade, a qual muitas vezes está relacionada à quantidade e à relevância de seus conteúdos acadêmicos. Outra faceta da análise de conteúdo também permite identificar o tema das discussões, e, com isso, acompanhar o progresso da aprendizagem (LUI et al 2004).

Gunawardena, Carabajal e Anderson (2001) propuseram um modelo de análise de interações para avaliar a construção do conhecimento em um grupo por intermédio da troca de informações em comunidades de aprendizagem online.

Swigger et al. (2010) compararam e avaliaram dois métodos de análise de conteúdo gerado por grupos de alunos envolvidos em um projeto de desenvolvimento de software. Eles aplicaram técnicas de análise de conteúdo e métodos de classificação semi-automática de texto às mensagens trocadas em um projeto global de software de estudantes envolvendo alunos de os EUA, Panamá e Turquia. Ambos os métodos foram avaliados em termos de capacidade de prever o desempenho da equipe. A fim de determinar a relação entre o desempenho e o padrão de comunicação estabelecido ao longo das tarefas, os autores classificaram as mensagens trocadas entre as equipes usando uma sequência de categorias desenvolvida por Curtis e Lawson (2001). Mais de 2896 transcrições foram codificadas e analisadas dessa maneira. Ao fim do trabalho, Swigger et al (2010) concluíram que equipes de alto desempenho desenvolvem padrões de comunicação consistentes que podem ser contrastados com os de equipes de menor desempenho.

Em geral, a análise de conteúdo ainda permanece sendo aplicada manualmente, o que dificulta seu emprego em larga escala por se tratar de uma tarefa intensamente trabalhosa e demorada. Demanda-se, portanto, aplicações automatizadas, que tornem a classificação eficiente, permitindo que o processamento de grandes volumes de informação seja realizado em minutos, ao invés de dias. Recentes avanços na área de recuperação da informação e de mineração de dados têm contribuído para atender a essa demanda (IKONOMAKIS, KOTSIANTIS, TAMPAKAS, 2005). A classificação automática de textos tem sido um método vital para lidar com a crescente quantidade de informações, especialmente no ensino e aprendizagem.

McLaren et al (2007) e Rosé et al (2008) empregaram classificadores automáticos para acompanhar o progresso das discussões através de diferentes dimensões, por exemplo, o pensamento crítico e a argumentação, como uma forma de dar suporte a alunos em fóruns eletrônicos e prover aos professores ferramentas que facilitem a moderação em diversas discussões simultâneas.

Lui, Li, Choy, (2007) estudaram o potencial e as limitações da classificação automática de textos para a avaliação, a pesquisa e o monitoramento do progresso do ensino e da aprendizagem. Uma tarefa essencial para a classificação automática é a representação adequada dos textos. Diferentes palavras ou frases podem ser consideradas, o que constitui

um problema para o desempenho dos algoritmos de classificação. Portanto, antes da aplicação desses algoritmos, é necessário reduzir a dimensão dos atributos usados para caracterizar os textos. Basicamente, essa redução ocorre por meio da seleção de um subconjunto menor de atributos ou pela criação de um novo conjunto a partir de transformações aplicadas ao conjunto original. Com isso, o desempenho de um classificador está relacionado às características de seu corpus. Por exemplo, uma palavra pode ter muitos significados (polissemia) e diferentes palavras também podem representar o mesmo significado (sinonímia). Somente através do contexto no qual cada palavra está inserida poderá se entender precisamente esses significados, o que, muitas vezes, não é algo tão simples a se fazer. Portanto, para se alcançar bons resultados, é preciso construir *corpora* adequados e de qualidade. Um tema de pesquisa bastante ativo, com muitos desafios a explorar, é o emprego de conjuntos de treinamento eficientes para aprimorar o desempenho de classificadores.

É certo que os resultados já alcançados descortinam o início de uma longa jornada a ser percorrida, mas, nem por isso, deixam de ser encorajadores. O conteúdo das discussões eletrônicas ainda impõem grandes desafios às técnicas existentes. As mensagens são geralmente curtas, o que reduz a possibilidade de ocorrências de atributos que distinguem uma categoria. Além disso, as discussões normalmente contêm vários erros e suas mensagens apresentam-se incompletas e desestruturadas. Mesmo assim, a classificação automática pode ser aplicada para alguns fins específicos, onde precisões elevadas não são exigidas. Por exemplo, a avaliação de alunos requer alta precisão (90% - 100%), uma vez que se trata de uma medição de desempenho escolar. Para realizar moderação em fóruns eletrônicos de discussões, uma taxa de acerto entre 70% a 90% é suficiente (MCLAREN et al, 2007; DONMEZ et al, 2005). Em contrapartida, com uma precisão em torno de 60%, um classificador automático pode ser integrado a fóruns eletrônicos de discussões como uma ferramenta de análise do progresso da aprendizagem (LUI, LI, CHOY, 2007).

1.7 ORGANIZAÇÃO DA PESQUISA

A dissertação encontra-se estruturada em seis capítulos, sendo o primeiro esta Introdução e o último a Conclusão.

No Capítulo 2, são apresentados os principais pontos que serviram de base para fundamentar essa pesquisa em três áreas distintas: tecnológica, pedagógica e filosófica.

O Capítulo 3 mostra a evolução do trabalho. Através de uma contextualização

circunstanciada, descreve as etapas realizadas a partir da caracterização do problema até a proposta de solução, passando pelos trabalhos iniciais, com suas idas e vindas, que nos fizeram concluir a dar um novo rumo para a pesquisa.

No Capítulo 4, descrevemos o desenvolvimento do Módulo de Avaliação e, no Capítulo 5, encerramos descrevendo o experimento realizado para verificar se a estratégia de classificação desenvolvida é capaz de garantir o desempenho necessário para dar suporte a esta pesquisa.

CAPÍTULO 2: REFERENCIAL TEÓRICO

2.1 OS CAMINHOS PARA O EMBASAMENTO TEÓRICO

Em uma realidade complexa, percebemos que diversas transformações estão sendo provocadas pelo avanço acelerado da tecnologia. Do ponto de vista educacional, essas transformações foram capazes de alterar o foco do processo de ensino e aprendizagem para as interações.

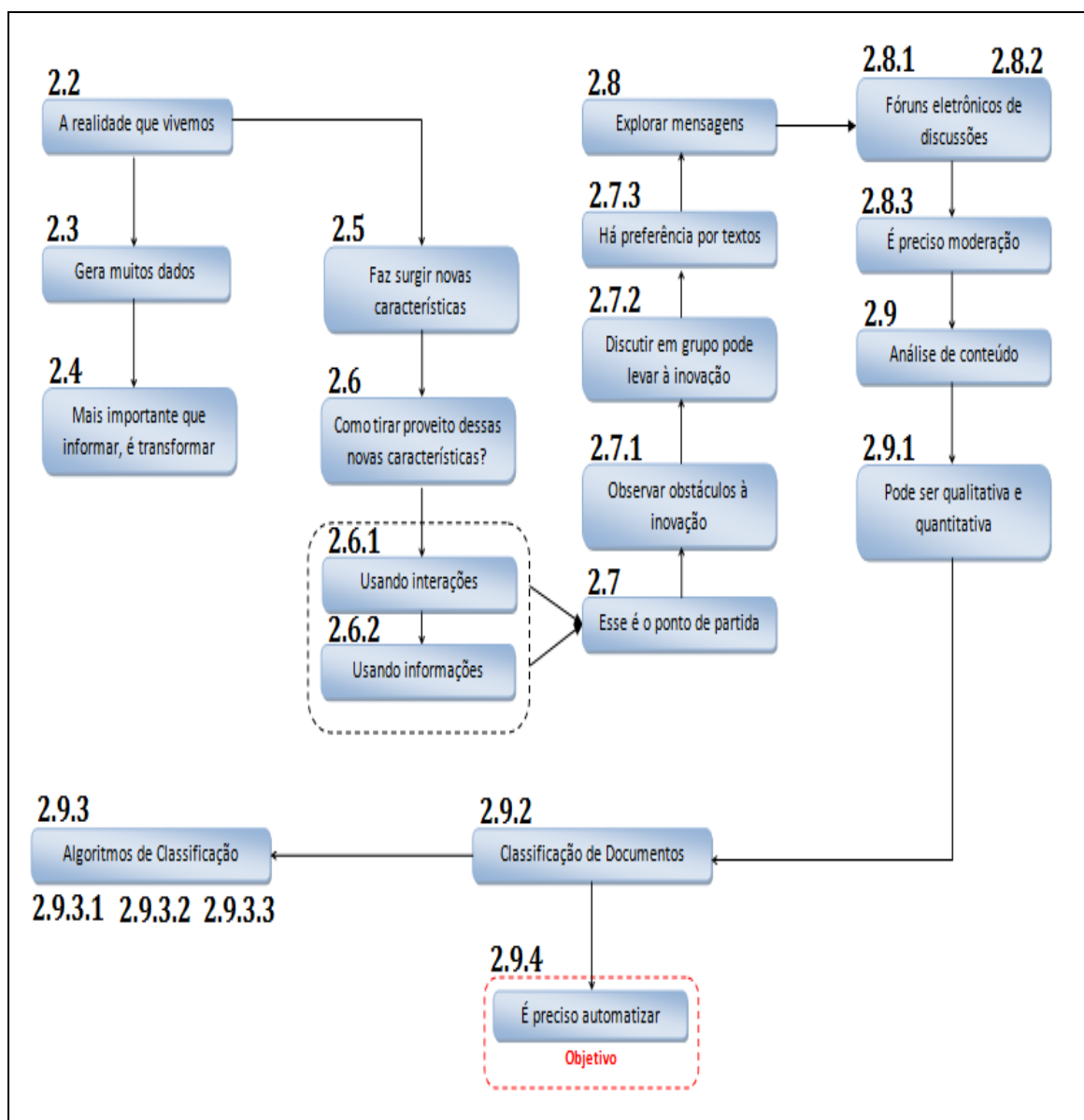


Figura 1. Referencial teórico

O diagrama da Figura 1 ilustra os caminhos que percorremos para tirar proveito das transformações que estão ocorrendo em favor do ensino e da aprendizagem. Nas próximas

seções, apresentamos com mais detalhes os questionamentos e as considerações que fundamentaram esses caminhos para nos levar até o objetivo desta pesquisa.

2.2 A REALIDADE QUE VIVEMOS

Presenciamos um enorme impacto provocado pelos meios de comunicação em termos de popularização do acesso à informação. Pessoas são "conectadas" de maneira que passam a sofrer fortes influências do meio em que vivem e do comportamento dos que os cercam. Estamos em um momento de epidemias sociais (GLADWELL, 2009), onde ideias se espalham como se fossem vírus, capazes de contagiar comportamentos e modificar tendências. Nossas experiências passam a ser sementes de transformações, para nós e outros também.

As demarcações dos conceitos, outrora bem nítidas, se tornam confusas rapidamente, sendo preciso testar repetidas vezes a forma como entendemos as coisas.

Não é um exagero dizer que a informação brota abundantemente por todos os cantos, já com o prazo de validade quase vencido. As questões em voga não se preocupam mais tanto em saber "onde está a informação", conforme antigamente. O problema passou a ser "o que" fazer enquanto ainda a informação é válida.

Com esse pano de fundo, pequenas mudanças são capazes de provocar efeitos extraordinários, a ponto de o mundo parecer não mais corresponder a nossa intuição. Mas, é necessário um estímulo certo. A informação, gerada e disseminada rapidamente, precisa convencer. Deve ser contextualizada para ser recebida em ambiente e momento adequados, capazes de garantir o entendimento das mudanças.

2.3 DADOS, DADOS E MAIS DADOS

Temos visto o mundo ser transformado por tecnologias que, em um curto prazo de tempo, conseguem atingir multidões de pessoas a custos reduzidos rapidamente. Cada vez mais e mais, usuários adotam e utilizam novos dispositivos e aplicações para interagir.

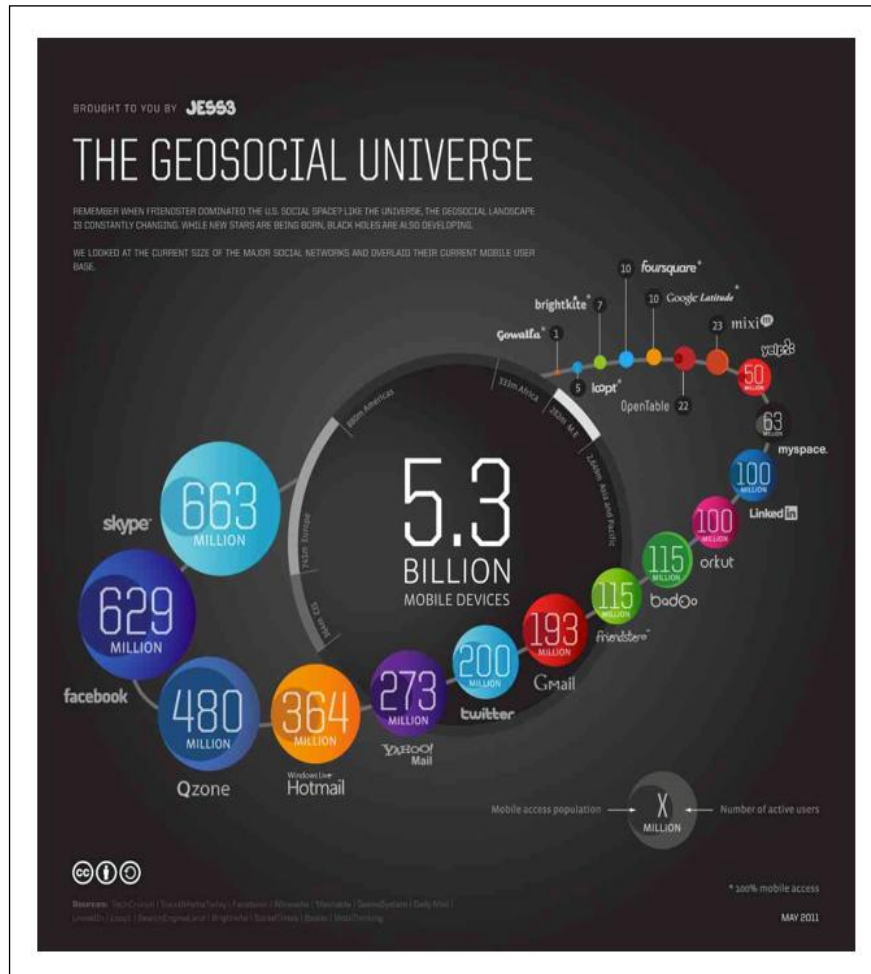


Figura 2. O universo geosocial da Internet. Fonte JESS3

Só os dispositivos móveis, tipo *smartphones* e *tablets*, já computam mais de 5,3 bilhões de exemplares, em um mundo que, há pouco, se aproximou da marca de 7 bilhões de habitantes³.

Tais números tornam-se ainda mais notáveis se considerarmos que, há 12 anos atrás, éramos 6 bilhões de pessoas, período que boa parte desses dispositivos ainda nem existiam ou não eram amplamente comercializados.

Se essas taxas se mantiverem, em breve, haverá mais dispositivos em rede do que pessoas no planeta. A estimativa é de que será o dobro da população mundial em 2015⁴.

O crescente volume de informações disponibilizadas a cada dia, que já equivale a milhares de anos de desenvolvimento da humanidade, é uma consequência direta da

³ Conforme Relatório sobre a Situação da População Mundial 2011 do Fundo de População das Nações Unidas (UNFPA/ONU). Disponível em:

http://www.unfpa.org.br/novo/index.php?option=com_content&view=article&id=795. Acessado em 31 dez 11.

⁴ Cisco Visual Networking Index: Previsão e Metodologia, 2010-2015. Acessado em 31 dez 2011. Disponível em: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html.

penetração efusiva da tecnologia na sociedade.

Nos últimos 5 anos, somente o tráfego IP (*Internet Protocol*) mundial anual aumentou oito vezes e a previsão para os próximos cinco anos é crescer quatro vezes mais. Até o final de 2015, deve atingir o limiar *Zettabyte* (ZB)⁵. Estima-se que será de 966 *Exabytes* (EB), o que equivale a 80,5 EB por mês. Para uma melhor compreensão do que esses números representam, 325 EB é aproximadamente a quantidade de dados que foi transmitida pela Internet desde a sua criação até o final de 2010⁶.

Impulsionado em parte pelo aumento da quantidade e da capacidade dos dispositivos, acredita-se que esse tráfego será de 11 GB per capita em 2015. No caso dos dispositivos móveis, a expectativa é de que ocorra uma elevação da proporção de tráfego na Internet, de 3% em 2010, para 15% em 2015, superando o volume gerado pelos dispositivos com fio.

Alguns fatos já reforçam essas estimativas, afastando a ideia de que sejam apenas opiniões futurologistas de mercado. Aplicações estão sendo migradas do desktop para a computação em nuvem e programas proprietários, tornando-se serviços abertos.

Enquanto isso, o crescente número de coletividades que agem colaborativamente tem originado um universo paralelo de comunidades virtuais. Diversos sites, que abrangem de mundos virtuais personalizados ao comércio social, oferecem diferentes tipos de recursos para conversação, o que tem modificado drasticamente o cenário das interações eletrônicas⁷ (Figura 3).

Levantamentos feitos pelo *Radicati Group*⁸ indicam que, em 2010, os 1,88 bilhões de usuários de correio eletrônico em todo o mundo enviaram quase 107 trilhões de mensagens. Foram cerca de 294 bilhões mensagens por dia ou, o que mais impressiona, 2,8 milhões por segundo. Em relação ao ano anterior, houve um aumento de aproximadamente 480 milhões de novos usuários. Projeta-se que até 2014 serão 3,8 bilhões de contas de correio eletrônico. É bem verdade que cerca de 89,1% dessas mensagens são spam e vírus. As mensagens

⁵ Um *Exabyte* (EB) equivale a 10⁹ Gigabytes (GB). Equivale a 250 milhões de DVDs. Com 0,0004 EB é possível armazenar a biblioteca digital de todos os livros já escritos, em qualquer idioma. Um *Zettabyte* é 1000 EB.

⁶ Minnesota Internet Traffic Studies (MINTS). University of Minnesota. Disponível em: http://www.dtc.umn.edu/mints/news/news_all.php. Acessado em 01 jan 2012.

⁷ Desde que foi produzido pela primeira vez em 2008, por Brian Solis, o infográfico *The Conversation Prism*, modificou bastante. A princípio, foram identificados 22 diferentes tipos de recursos para conversação. Já a versão mais atualizada disponível, a de 2010 (3ª versão), tem nada menos que 28 tipos diferentes, com variações dentro de cada um deles. Disponível em: <http://jess3.com/the-conversation-prism-v3/>.

⁸ Em: *Email user numbers and counts from Radicati Group (the number of sent emails was their prediction for 2010, so it's very much an estimate)*.

ainda mais.

Apesar de ainda haver algumas distorções entre suas regiões, o Brasil não fica longe dessas tendências globais. De acordo com levantamentos da Agência Nacional de Telecomunicações (ANATEL)⁹, de 2007 a 2011 houve uma expansão em todas as modalidades de serviços do setor telecomunicações. São mais de 202 milhões de linhas de telefonia móvel celular. Em novembro de 2010, esse número já havia superado o da população do país¹⁰.

No Serviço de Comunicação Multimídia (SCM), responsável por dar suporte para o acesso à Internet, a cada 100 domicílios, pelo menos 26 possuem um ponto de acesso fixo. Esse é um dos indicadores que serve para mostrar o crescimento do número de usuários ativos da Internet no Brasil, principalmente nas residências¹¹.

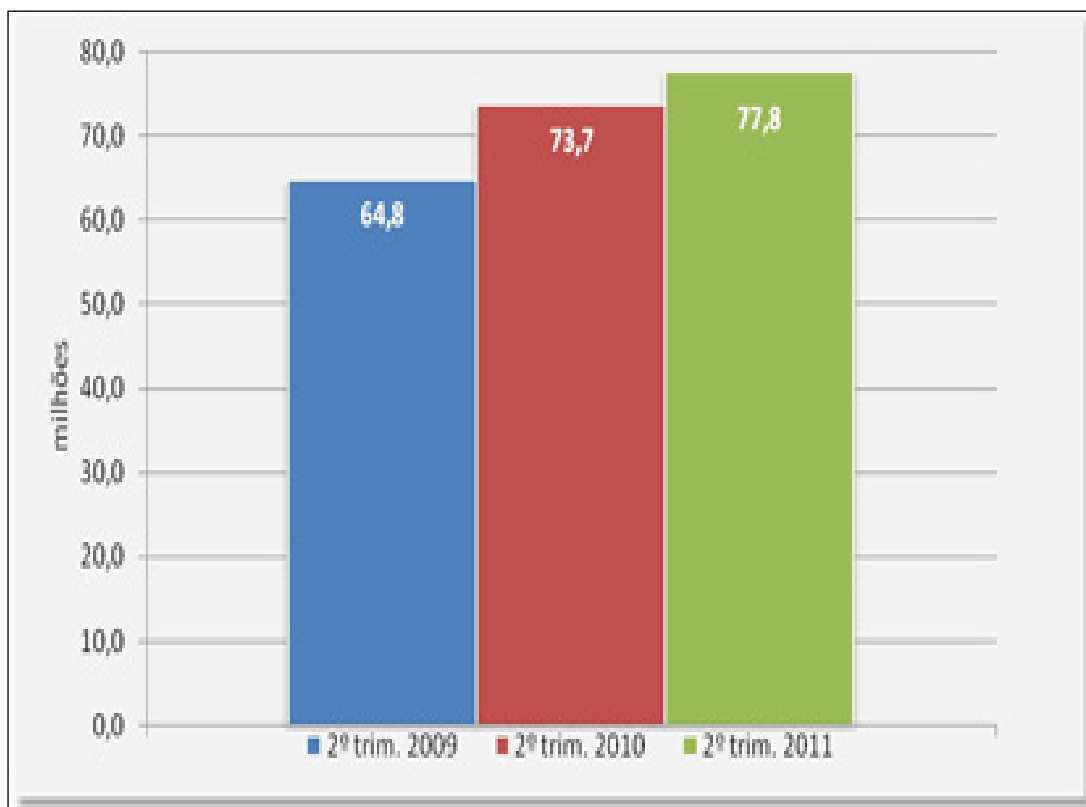


Figura 4. Evolução do número de pessoas com acesso à Internet no Brasil. Fonte: NetView – IBOPE Nielsen Online

⁹Em: ANATEL - Números do Setor de Telecomunicações no Brasil - 2007 a 2011, Disponível em: www.anatel.gov.br. Acessado em 29 dez 11.

¹⁰De acordo com dados do IBGE e da ANATEL. Veja em <http://oglobo.globo.com/economia/numero-de-celulares-no-brasil-maior-que-de-habitantes-2924116#ixzz1hwx5hkZF>

¹¹De acordo com a Ibope Nielsen. Disponível em:

<http://www.ibope.com.br/calandraWeb/servlet/CalandraRedirect?temp=5&proj=PortalIBOPE&pub=T&db=cald&comp=Noticias&docid=C2A2CAE41B62E75E83257907000EC04F>

Apenas no segundo trimestre de 2011, 77,8 milhões de usuários utilizaram a Internet, seja em casa, trabalho ou escola, representando um aumento de 14% em relação ao mesmo período do ano passado. Com este resultado, o Brasil ficou atrás apenas de Japão e Estados Unidos.

O barateamento dos equipamentos de informática e a inclusão digital da nova classe C são considerados dois fatores que contribuíram para esses resultados. Estima-se que, para o próximo ano, esse avanço continue ainda maior.

Entre as categorias de sites mais visitados está a de Educação, com um crescimento percentual, em agosto de 2011, de 9,1%, relativo ao mês anterior, atingindo 25,8 milhões de usuários. Nesse mesmo período, 87% dos internautas brasileiros usaram sites sociais. No total, foram 39,3 milhões de pessoas que acessaram redes sociais, fóruns, blogs, *microblogs* e outras páginas de relacionamento. Somente o *Facebook* foi acessado por 30,9 milhões de usuários; o *Orkut*, por 29 milhões; o *Twitter*, por 14,2 milhões. Na média, cada um desses usuários conectou-se por um tempo de 7 horas e 14 minutos. Esses resultados posicionam o Brasil entre os países com maiores alcances em sites sociais e indicam que a nova tendência mundial, de criar e manter relacionamentos virtuais, também é uma realidade entre nós.

2.4 MAIS IMPORTANTE QUE INFORMAR, É TRANSFORMAR

Antigamente, as informações não eram tão atualizadas e precisas, o que exigia muito esforço para sua compreensão. Por exemplo, era comum irmos a uma biblioteca para fazer pesquisas sobre determinado assunto.

Hoje, fazemos a biblioteca (*Wikipédia*) e ainda assim, a falta de discernimento é alta. Informação farta e disponível rapidamente com facilidade, ao mesmo tempo, estimula o "ser construtivo", que busca por novos caminhos e o "ter descartável", que gera desinteresse rápido. Normalmente, ficamos com a segunda opção, a mais fácil. Dessa maneira, a abundância de informação pode desvalorizar a busca pelo conhecimento e gerar ensinamentos descartáveis, que produzam ações mecânicas de copiar e colar. Buscar atalhos e não caminhos para uma solução legítima é o que se aprende com isso. E, o que se alcança são dificuldades, como falta de concentração e atenção.

Oferecer informações não significa que conhecimento será gerado. Isso depende de como, quando e onde as informações serão usadas. Temos tido informações de mais para tempo de menos. Tudo é muito rápido e corrido, o que tem levado a transtornos como

depressão, stress e ansiedade lamentavelmente, cada vez mais comuns em nossa sociedade.

É preciso que o aluno seja desafiado a ir além do "copiar e colar". Deve percorrer caminhos que, ainda que mais longos, levem à contextualização e compreensão dos fatos.

Toda a discussão em torno dos efeitos provocados pela tecnologia não é sobre as ferramentas em si, mas sobre seu uso. Lidar com os anseios específicos de uma geração, que cada vez mais valoriza a informação, sem deixar de levar alunos a pensar por conta própria e a manter o foco para comunicar suas ideias com clareza, garantindo a credibilidade do processo de ensino e aprendizagem, trata-se de uma questão complexa e um desafio para as instituições de educação (BARNES, 2007).

A despeito de ainda não parecer ser uma unanimidade, consideramos que o emprego da tecnologia pode ser bom para a sala de aula, desde que aumente o que já está lá em termos de interação, colaboração e comunicação (BARNES, 2007).

Não há dúvidas que as características da nova geração de estudantes representam um desafio às práticas educacionais existentes. Experiências de auto-descoberta proporcionadas pela forte exposição à Internet e a saturação de outros recursos tecnológicos proporcionaram formas distintas de pensar, comunicar e aprender ainda não encontradas em gerações anteriores (PRENSKY, 2006; TAPSCOTT, 2010).

Assim, tecnologias que aumentam a autonomia na aprendizagem conseguem alterar o sistema educacional. Todavia, a forma de empregá-las determina se para melhor ou para pior.

Se forem usadas separadamente dos processos da vida dos aprendizes, serão ferramentas pedagógicas inadequadas (PAPERT, 1993). E, para aqueles que "cresceram digitalmente," a realidade é a da interação social. As relações com os colegas representam algo muito importante para o dia a dia, servindo inclusive como estímulo para aprender.

Mas, não para por aí. É uma questão que vai muito além da ambientação ou contextualização apenas, que extrapola o choque entre gerações. A tecnologia é uma realidade para todos nós. Favorável para alguns, nem tanto para outros, é preciso saber tirar proveito desta situação. Em tese, a nova realidade tem significados diferentes para os aprendizes, mas, na prática, sustenta-se por expectativas comuns a todos por abordagens educacionais inteligentes, que enfatizam interação e colaboração, motivadas pelo próprio avanço tecnológico. De igual maneira, desejamos ser envolvidos por práticas que trazem significado à aprendizagem e que considerem nossas preferências. Portanto, ainda que as características dos alunos da "geração Net" estimule estratégias pedagógicas que considerem o uso de recursos

tecnológicos para a aprendizagem, é um desperdício tratar deste fato apenas como um diferencial que beneficia uma geração e desfavorece outra.

Preferimos enxergar a tecnologia como um instrumento poderoso para o desenvolvimento de argumentação e inovação, características essenciais a serem realçadas em todos nós. Acreditamos que empregar recursos tecnológicos simplesmente para motivar alunos, dando-lhes a oportunidade de usar equipamentos com os quais estão acostumados, ou não, a interagir no cotidiano, não é um procedimento tão eficaz para a aprendizagem quanto estimulá-los a desenvolver a compreensão e o pensamento crítico para as diferentes fontes de informações que surgem a cada momento.

O ponto-chave é levá-los a selecionar e avaliar continuamente as informações por meio de questionamentos próprios, até que consigam reunir evidências para sustentar suas respostas e criar novas ideias (HOWARD, 2006). Temos por certo que, se aplicada adequadamente, a tecnologia pode proporcionar interações que nos capacitarão para nos adaptarmos às situações inéditas e resolver os problemas da vida.

2.5 NOVAS CARACTERÍSTICAS DOS APRENDIZES

Os aprendizes estão saturados de meios de comunicação. A facilidade para se comunicar trouxe o acesso rápido à informação, criando uma cultura de imediatismo, como se tudo estivesse "*online*", a um clique de acontecer. Essa impressão rouba o exercício da paciência, embaça a clareza do significado das experiências e torna os aprendizes menos tolerantes a atrasos, desejosos por respostas rápidas e por um ensino no qual não precisam mais esperar aprender para, depois, praticar, e sim, aprender já praticando.

Outro aspecto da diversidade de meios de comunicação que serviu para dar nova característica aos aprendizes é o que chamamos de "paradoxo tecnológico". Uma das grandes expectativas pelo avanço da tecnologia sempre foi a capacidade de disponibilizar mais oportunidades para que se pudesse gozar livremente o tempo. No entanto, o que se tem observado é um efeito contrário. A falta de tempo tem obrigado os aprendizes a cada vez mais se acostumarem com múltiplos estímulos e se esforçarem para fazer várias tarefas simultaneamente (OSER, 2005). Como consequência direta desse "efeito multitarefa", estão mais apressados, permanecem menos tempo atentos às atividades e deixam de fazer a triagem das informações com que lidam. Acumulam conhecimentos que se aplicam às suas necessidades e logo os descartam. A sensibilidade aos fatos mais relevantes depende do

quanto estão envolvidos e motivados. Correlacionam coincidências para dar sentido as suas respostas. Ocupados com o presente, tendem a ignorar o passado e a se despreocupar com o futuro. Preferem fazer rápido a fazer bem feito.

Os aprendizes também são pensantes, criativos e imaginativos. Já trazem consigo uma bagagem cheia de conhecimentos, mas superestimam o que já sabem (RUSSO, SCHOEMAKER, 2001). Encontram dificuldades para associar conhecimentos antigos a novos, assim como contextualizar e criar hipóteses.

Os aprendizes estão direcionados para objetivos. Uma vez que tempo é precioso, os aprendizes rejeitam atividades que não sejam relacionadas aos seus objetivos. Dificilmente "absorvem" aquilo que deixa de satisfazer suas expectativas (HAY, 2000). Apesar de terem habilidades para se adaptarem e espírito explorador, precisam de estímulo para buscar novas oportunidades de soluções em caminhos mais longos, porém mais compensadores. Preferem percorrer caminhos mais curtos para chegarem logo a uma solução, mesmo que seja trivial. Buscam atividades direcionadas em ambientes interativos, que lhes envolvam no processo de aprendizagem e que lhes proporcionem experiências com significado (GLENN, 2000). Querem *feedback* de suas ações. Guardam sucessos, mas desperdiçam as oportunidades para aprender com os erros. Ficam inseguros e duvidosos com maus resultados e tentam esquecê-los rapidamente.

Os aprendizes desejam autonomia e independência. Com tantas fontes de informações disponíveis, os aprendizes direcionam suas escolhas para "o que", "onde", "como" e "com quem" aprender, selecionando apenas aquelas que empregam técnicas de ensino mais adequadas as suas preferências e que criam experiências com significados (BARNES, 2007).

Os aprendizes querem o coletivo, mas gostam de exclusividade e personalização. A profusão de redes sociais, principalmente entre os jovens, é um indicativo de que têm grande expectativa por trabalhos em grupo e projetos de pesquisa interativos. Por outro lado, os inúmeros apetrechos que utilizam para personalizar suas páginas e seus perfis é um indicativo de que desejam mostrar quem são. Gostam de ter por opções que atendam suas preferências. Também gostam de tratamentos diferenciados. Esperam ser reconhecidos e compensados, principalmente quando são bem sucedidos.

2.6 COMO TIRAR PROVEITO DAS NOVAS CARACTERÍSTICAS DOS APRENDIZES

O desenvolvimento tecnológico foi uma semente de transformação, que trouxe rapidez e novas formas para lidar com a informação e interagir. Mudamos de agentes passivos para ativos. Deixamos de apenas correr atrás da informação, para absorvê-la e armazená-la como é, e passamos a transformá-la de fato.

A tecnologia integrou e criou meios de comunicações, que nos tornaram "conectados". Ganhamos facilidades para compartilhar ideias, algo que, outrora, era restrito a alguns.

Recursos cada vez mais potentes têm sido disponibilizados em massa para facilitar o acesso à informação, acumulando um enorme volume de dados, que são alterados a todo o momento.

Com tanta informação assim disponível, manter o foco, sem perder o rumo, é uma tarefa árdua. Nossa atenção é requisitada o tempo todo e não dá para seguir em frente sem eliminar a sobrecarga cognitiva e certificar a validade das informações. Depender apenas do acaso para descobrir novos caminhos, muito menos. Para ser produtivo na sociedade do conhecimento, é essencial fazer uso significativo da informação.

Inovar é a palavra de ordem que tem conduzido à construção do que poderíamos chamar de "sociedade da inovação". Para tanto, é preciso, oportunamente, saber distinguir o que é válido, selecionar o que é importante e modificar o que é comum. Dar conta dessas tarefas significa ter habilidades para analisar, compreender, avaliar, criticar, argumentar, classificar, organizar e criar.

Do ponto de vista pedagógico, significa que, em um mundo cheio de complexidades, onde as informações mudam o tempo todo, o ensino e a aprendizagem também devem ocorrer constantemente. Por isso, é necessário que sejam: simples, para serem repetidos o quanto necessário; precisos, para não perder boas oportunidades; eficientes, para trazer resultados que motivem a seguir em frente.

2.6.1 USANDO INFORMAÇÕES

Através da própria tecnologia é possível empregar recursos que tragam as características necessárias para o processo de ensino e aprendizagem continuado. O carro-

chefe é fazer uso da imensa riqueza de informações, como base para criar mecanismos customizados, capazes não apenas de satisfazer anseios individuais, mas também de proporcionar ações que aperfeiçoem características e desenvolvam as habilidades necessárias. Algumas dessas ações são:

- promover a "aprendizagem do aprender": investigar, entender e transformar;
- incentivar ações criativas de professores e alunos;
- favorecer o intercâmbio dos saberes;
- provocar desequilíbrio entre perguntas e respostas, para aumentar a produtividade;
- desenvolver a construção de novos conhecimentos através do pensamento crítico;
- estimular o "aprender fazendo" (BARNES, 2007; PRENSKY 2006), para reduzir a distração e promover autonomia;
- trabalhar conhecimentos já aprendidos. Erros e acertos passados podem levar a um mapa mais confiável da trajetória do futuro.
- aproveitar o interesse pela potencialização do coletivo, para o crescimento individual.

2.6.2 USANDO INTERAÇÕES

A tecnologia pode contribuir para que nos relacionemos com as pessoas certas, que nos farão aprender mais e melhor.

A disseminação de ideias depende mais da estrutura das redes de contatos do que das características individuais das pessoas (CHRISTAKIS, FOWLER, 2010). Quando não temos certeza sobre como devemos agir, costumamos observar as pessoas ao nosso redor. Isso ocorre porque ao observarmos os outros, nossos cérebros são capazes de simular o que eles estão sentindo (KEYSERS, 2009). Mas, nem sempre esse efeito é consciente. É possível mudarmos nosso comportamento com base naquilo que as pessoas que nos rodeiam estão fazendo, sem percebermos que estamos sendo influenciados por eles. Por exemplo, aprendizes que têm companheiros estudiosos tendem a estudar mais (KEYSERS, 2009).

Apesar de recebermos influências de uma grande quantidade de pessoas, aquelas que percebemos serem parecidas conosco são as que mais nos influenciam. Esse efeito se torna ainda maior quando nos comparamos com outras pessoas semelhantes a nós (FOGG, 2003).

Olhar também toda hora para fora nos faz perder nossos desejos, histórias, objetivos, autoconhecimento. Nos tornamos pessoas mais indecisas e inseguras. Se não aprendermos a ficar solitários, seremos sempre solitários. Isso não significa que devemos pensar em nos desconectarmos. Não temos essa opção. O que precisamos é descobrir como fazer opções saudáveis para nos conectarmos com qualidade. Fazer regularmente o "*upload*" de nossas próprias ideias e nutrir-las para que permaneçam ativas e interativas nos torna melhor, e outros também.

2.7 O PONTO DE PARTIDA

Presenciamos o momento da "tecnologia persuasiva", em que sistemas de comunicação interativos são desenvolvidos para influenciar as atitudes e comportamentos das pessoas (FOGG, 2003). Vemos reunido na comunicação o binômio "interação-informação", apropriado para se alcançar inovação.

Com isso, decidimos que o ponto de partida de nosso trabalho deveria ser procurar por um mecanismo de comunicação adequado, capaz de promover ensino e aprendizagem inovadores.

2.7.1 OBSTÁCULOS À INOVAÇÃO

A inovação resulta de processos sociais. Apoia-se no trabalho em equipe e na colaboração. Quando pessoas trabalham juntas em busca de um objetivo comum, ela flui com mais facilidade. Mas para isso, exige pensar diferente, através de pontos de vista que normalmente não são considerados. Também depende da comunicação para que sejam desenvolvidas ideias e soluções alternativas. Alguns dos principais obstáculos para o desenvolvimento da inovação em um grupo são (THE MCGRAW HILL SMALL GROUP COMMUNICATIONS, 2011):

- **Pressão para se conformar às normas do grupo.** Sem normas, os grupos não são capazes de funcionar. No entanto, a preocupação demasiada e intolerante com o cumprimento dessas normas sufoca a criatividade e torna a inovação algo difícil de ser alcançado. Primar pela unanimidade, em vez de incentivar a apresentação de perspectivas legítimas, inibe o surgimento de novas ideias.
- **Incapacidade ou falta de vontade para colaborar.** Ideias criativas e inovadoras

raramente ocorrem quando há um clima exagerado de competição entre os membros do grupo. A desconfiança aumenta e, em vez de compartilhar as informações, ficam propensos a retê-las. Com isso, acabam se concentrando em si mesmos e esquecem os objetivos do grupo. As vezes, os membros não possuem as competências necessárias para promover a colaboração.

- **Comunicação defensiva ou insuficiente.** Quando os membros estão inseguros quanto a seus parceiros, eles interagem defensivamente ou quase não se comunicam, deixando pouco espaço para novas ideias. Receosos com avaliações críticas descabidas e desmotivadoras tendem a não apresentar seus pensamentos e, com isso, ideias potencialmente úteis são perdidas.
- **Diferenças de estilos.** Por mais que os membros sejam parecidos, sempre há características que os permitem distinguir. Tais diferenças definem a maneira como cada um deles participará do grupo. Por exemplo, alguns são mais sinceros, extrovertidos e falantes para expressar suas opiniões, enquanto outros são mais reservados, introvertidos e hesitam em se pronunciar. Naturalmente, os que têm mais facilidade para se comunicar dominarão as conversações e suas ideias serão as mais ouvidas. Como consequência, tenderão a liderar e influenciar os demais. Por sua vez, membros mais reticentes podem ser intimidados pelos ríspidos ou truculentos a não acrescentar suas opiniões às discussões.
- **Aspectos culturais.** Criatividade e inovação surgem quando pessoas vêem o mundo de maneira única. Mas, normas culturais, valores e crenças podem estar tão enraizados que impedem isso de acontecer. Quando são inflexíveis e intolerantes, forçam os criativos a serem considerados como revolucionários não conformistas, trazendo descréditos às suas ideias inovadoras.

2.7.2 DISCUTIR EM GRUPO PODE LEVAR À INOVAÇÃO

Quando se age em grupo, pode-se alcançar melhores resultados do que sozinho. Decisões mais precisas, ideias mais criativas e soluções inéditas costumam aparecer mais vezes. Porém, apenas sob certas condições, isso acontece (ADAMS, 2011).

Normalmente, ocorre quando os grupos são bem ajustados, compostos por integrantes ativos, que discutem a respeito de suas convicções e insinuações (BAHRAMI et al, 2010) e

que demandam explicações e justificativas de seus parceiros (SOLLER et al, 1999) antes de tomarem uma decisão.

É necessário que haja diversidade de opiniões e liberdade para que as pessoas tomem suas decisões independentemente do ponto de vista daqueles ao seu redor. Cada membro também deve ser capaz de extrair conhecimento de forma descentralizada e estar motivado a compartilhá-lo com os demais companheiros. Por fim, as opiniões de todos precisam ser agregadas adequadamente para se tornarem decisões coletivas (SUROWIECKI, 2004).

2.7.3 PREFERÊNCIA PELO TEXTO

Apesar de estarmos testemunhando uma evolução de tirar o fôlego, provocada pelas novas tecnologias e com diversos recursos multimídias disponíveis, o texto tem se mantido como a forma preferida para se comunicar (Figura 5).

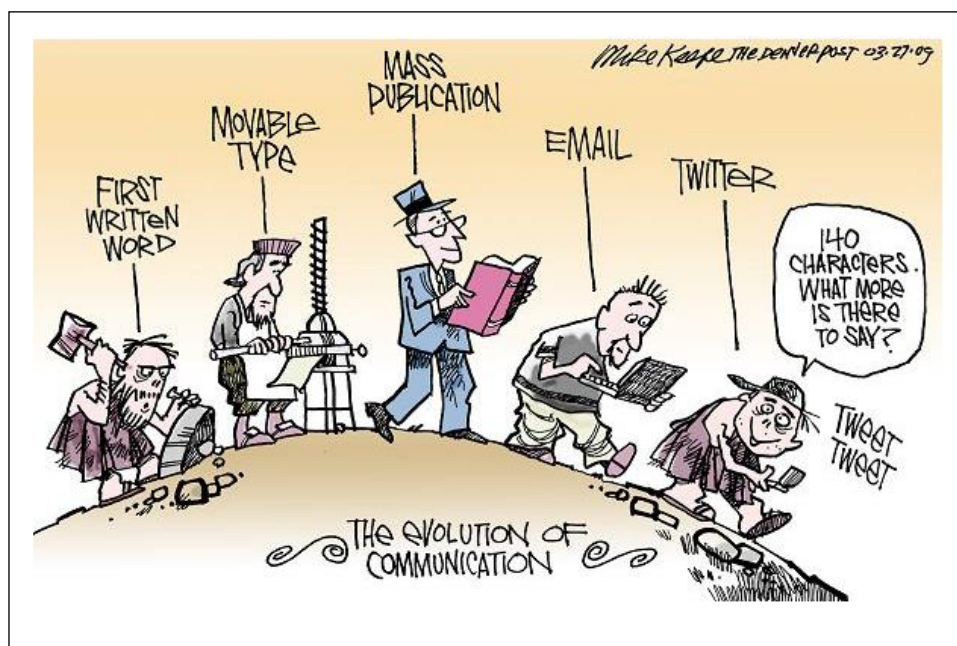


Figura 5. A evolução da comunicação. Fonte: Mike Keefe. Disponível em: <http://www.intoon.com/cartoons.cfm/id/68559>

As pessoas sempre buscaram novas formas de expressão rápidas e funcionais, que economizasse tempo e dinheiro.

Novas tecnologias, particularmente a Internet, têm oferecido comunicação, síncrona e assíncrona, em escala mundial, a um custo bem baixo.

Mas, parece haver uma preferência de não encarar a complexidade de uma conversa

falada. Em contrapartida, a comunicação escrita abafa emoções. É difícil fazer o sentimento ir por completo com as mensagens de um texto.

Talvez seja por isso que as pessoas têm cada vez mais optado por trocar mensagens em grupo, como uma forma de compensar essa dificuldade. A dinâmica emocional é muito mais forte do que sozinho. Isso explicaria uma tendência de transição do e-mail para as redes sociais, passando pelos *blogs*, chats e *wikis*.

2.8 EXPLORANDO AS INTERAÇÕES ESCRITAS: AS RIQUEZAS LATENTES DAS MENSAGENS

Na discussão assíncrona promovida pelos fóruns eletrônicos, as mensagens podem ser lidas a qualquer momento, pelos membros do grupo. Com isso, ganham mais tempo para refletir sobre o que os outros escreveram e pensar sobre o que dizer, para compor, refletir e editar suas próprias mensagens. A desvantagem é que os membros podem deixar de visitar o fórum regularmente, se não forem motivados.

Como as discussões são temáticas e armazenadas sequencialmente, além de se ter mais facilidade para seguir as mensagens do que por email, também é possível explorar essa grande quantidade de dados para detectar padrões ou regras de associações que permitam entender o processo de aprendizagem e também descobrir novas informações (SOLLER, LESGOLD, 2000).

Os padrões de conversações que um grupo utiliza podem indicar a qualidade de suas interações e, conseqüentemente, o resultado de seu trabalho. Conversações de grupos efetivos incluem uma mistura equilibrada de diferentes padrões, com uma particular abundância de questionamentos, explicações e motivações. Grupos não efetivos apresentam uma distribuição menos equilibrada dos padrões de conversação, com uma tendência acentuada de expressões socializantes (SOLLER et al, 1999). Em geral, grupos que apresentam mensagens relacionadas a comportamentos colaborativos tendem a apresentar desempenho mais elevado (SERCE et al, 2011).

Assim, através da análise das mensagens, é possível verificar a qualidade das interações escritas (SWIGGER et al, 2010; MACDONALD, 2003). Alunos poderão fazer uso dessa análise para refletir sobre suas performances; professores, para mediar e avaliar (LOPES, 2007); pesquisadores, como dados para identificação e análise de padrões de discussões (MEYER, 2004).

Ao longo da última década, diversos trabalhos, que tiveram como objetivo verificar o quanto e como a qualidade das interações escritas pode impactar a construção do conhecimento, foram beneficiados com farto material para ser analisado produzido em fóruns eletrônicos (LAZONDER, WILHELM, OOTES, 2003; PENA-SHAFF, NICHOLLS, 2004; CORICH, KINSHUK, HUNT, 2006; KULJIS, LINES, 2007).

2.8.1 FÓRUNS ELETRÔNICOS DE DISCUSSÕES: AMBIENTES CAPAZES DE PROMOVER O PENSAMENTO CRÍTICO

Presenciamos a popularidade e a aceitação dos fóruns eletrônicos de discussões como uma ferramenta de aprendizagem. Boa parte do conhecimento tem sido compartilhada livremente nos fóruns de discussão, criados em torno dos *blogs*, *chats*, *wikis* e redes sociais da Internet.

Os fóruns eletrônicos de discussões conseguem combinar as vantagens da comunicação mediada por computador (do inglês, *Computer Mediated Communications* - CMC) com as da sala de aula tradicional, onde alunos e professores não somente compartilham um espaço físico, mas também um conjunto de normas comportamentais e contextos comuns (ASTERHAN, 2010). Por isso, têm sido tipicamente usados para dar apoio a atividades de EAD, como a realização de palestras e seminários em cursos a distância pela Internet. Nesse contexto, as discussões assíncronas são as mais comuns. Mensagens assíncronas apresentam vantagens em relação às síncronas. Os alunos ganham mais tempo para procurar por informações adicionais e para refletir melhor antes de interagirem, o que lhes proporcionam discussões mais consistentes.

Em uma visão mais ampla, os fóruns eletrônicos de discussões não apenas disponibilizam recursos para estruturar e registrar o debate sobre um determinado assunto. Pesquisas têm mostrado que seu uso, tanto nas salas de aula, quanto em ambientes de EAD, podem oferecer outras vantagens, especialmente em relação aos aspectos sócio-interativos das discussões (ASTERHAN, EISENMANN, 2009).

Diferentemente de uma discussão em sala de aula, nos fóruns eletrônicos, os participantes sentem-se mais confiáveis, e confortáveis, para expressar livremente suas ideias, questionar, argumentar e aprender uns com os outros, sem interrupções e influências externas. Por isso, tendem a interagir mais com o grupo (ASTERHAN, 2010). E mais, encorajam outros a argumentarem também; a questionar princípios; a explicar e justificar opiniões; a articular, elaborar e refletir sobre seus conhecimentos, para chegar a uma conclusão. Com

isso, quanto mais argumentam, proporcionam discussões de maior qualidade, avançando na construção do conhecimento (LAI, 2008).

Assim, os fóruns eletrônicos de discussões também constituem um ambiente compartilhado capaz de promover interações que levam ao desenvolvimento da colaboração e do pensamento crítico dos alunos, ingredientes essenciais para a inovação (SCHEUER, MCLAREN, 2008).

2.8.2 FÓRUNS ELETRÔNCOS DE DISCUSSÕES: AMBIENTES PROPÍCIOS PARA AVALIAÇÃO

A participação e a integração social têm sido apontadas como responsáveis pelos bons resultados em EAD (BARILLI, 2006). Quando as pessoas são observadas em situações que envolvem relações sociais, suas ações são avaliadas com mais facilidade.

Nesse contexto, os fóruns eletrônicos destacam-se por proporcionarem contato com outras realidades sociais através da troca de ideias e debates (CAMPOS et al, 2003). Constituem ambientes onde ocorrem interação e colaboração, permitindo que sejam realizadas análises contínuas, que capturam erros e acertos durante o progresso do ensino. Assim, professores têm acesso aos efeitos de seu trabalho enquanto alunos podem identificar deficiências, reconhecer e corrigir os próprios erros. Com isso, estratégias podem ser revisadas e falhas corrigidas em tempo hábil.

Os fóruns eletrônicos também estimulam a construção de conhecimento, levando ao desenvolvimento individual e coletivo conforme necessidades específicas.

Com o diálogo sendo a essência da avaliação, ela se torna mais abrangente (NUNES, 2006). Ao contarmos nossas práticas, expomos as ideias que nos cercam e damos chances para que outros interpretem criticamente nossos posicionamentos, criando oportunidades de discussões que nos exigem defender pensamentos, analisar procedimentos, rever conceitos e reconhecer dificuldades.

Por tudo isso, consideramos os fóruns eletrônicos ambientes propícios para levantar informações úteis à avaliação do processo de ensino-aprendizagem.

2.8.3 É PRECISO MODERAÇÃO

Os benefícios proporcionados pela aprendizagem em grupo dependem da qualidade das discussões ocorridas (ASTERHAN, 2010). Apesar dos fóruns de discussão contribuírem para o desenvolvimento do pensamento crítico, verifica-se que muitas mensagens são infrutíferas. Muitas vezes, os alunos utilizam inadequadamente as ferramentas para conversas paralelas ao invés de permanecerem focados no tópico da discussão; geram contribuições sem sentido; ignoram argumentos e questões de outros colegas. Alguns chegam a permanecer isolados, sem participação ativa, enquanto que outros, dominam as discussões por completo. Por tudo isso, há a necessidade de ajuda e guiamiento.

O professor pode tirar proveito de sua experiência para aprimorar o debate, dando orientações quando ocorrerem problemas e provendo encorajamento quando as discussões forem produtivas (KING, ROSENSHINE, 1993; HAKKARAINEN, 1998; GILLIES, 2003; TOLMIE et al, 2005; WEBB, 2009). Entretanto, torna-se difícil para um único professor acompanhar na íntegra um debate em que diversas discussões ocorrem simultaneamente.

A tecnologia pode auxiliá-lo a realizar a mediação adequada. Através do emprego de ferramentas computacionais, capazes de processar, agregar e sumarizar o intenso fluxo de informações gerado pelas discussões, pode ser direcionado para os principais pontos discutidos e medir a qualidade das interações, para que as ações necessárias sejam tomadas oportunamente.

2.9 ANÁLISE DE CONTEÚDO

A análise de conteúdo pode ser definida como uma técnica de pesquisa que emprega procedimentos para fazer inferências em textos (ANDERSON, ROURKE, GARRISON, ARCHER, 2001).

Ela tem sido bastante utilizada para prover evidências acerca da aprendizagem e da construção do conhecimento em fóruns de discussões.

Em um primeiro momento, as pesquisas a respeito das discussões eletrônicas ficaram restritas apenas à coleta de dados quantitativos sobre os níveis de participação (HENRI, 1992). Contudo, esses índices quantitativos não serviam para determinar a qualidade das interações (MEYER, 2004).

Ao longo do tempo, a análise de conteúdo foi sendo adotada em larga escala como

uma técnica para detectar as informações que está incrustada nas mensagens dos fóruns eletrônicos de discussões. Henri chega a chamar a conferência mediada por computador (do inglês, *Computer Mediated Conferencing*) de "mina de ouro" das estratégias de aprendizagem (HENRI, 1992).

Outros pesquisadores usaram as transcrições das discussões eletrônicas para investigar o processo da construção social do conhecimento (GUNAWARDENA, CARABAJAL, LOWE, 2001; GUNAWARDENA, LOWE, ANDERSON, 1997) e do pensamento crítico (BULLEN, 1997; NEWMAN, WEBB, COCHRANE, 1995).

Apesar dessa técnica de pesquisa ser bastante usada, ainda não foram estabelecidos padrões para ela. Os diferentes instrumentos utilizados refletem uma variedade de abordagens e diferem no nível de detalhes e nos tipos de categorias empregadas. Há ainda diferenças adicionais relacionadas à diversidade das bases teóricas, à validade, à confiabilidade e à escolha para a unidade de análise (DE WEVER et al, 2006).

2.9.1 ANÁLISES QUANTITATIVA E QUALITATIVA

A comunicação exerce um papel fundamental no desempenho do trabalho em grupo e, por isso, tem sido o foco de diferentes pesquisas. Alguns dos resultados já alcançados mostram que, através da análise de interações por meio de mensagens trocadas assincronamente em fóruns de discussão, é possível avaliar e influenciar o desempenho do trabalho de um grupo (SWIGGER et al, 2010).

O método mais simples, e também comum, para realizar esse tipo de análise é o quantitativo, que se baseia na contagem de medidas de baixo nível (*low-level*), como o tamanho e o número de mensagens trocadas. Contudo, a análise quantitativa desperdiça informações importantíssimas, pois não consegue capturar aspectos relacionados ao intuito das mensagens, o que, muitas vezes, leva a resultados equivocados (SWIGGER et al, 2010).

A análise qualitativa, ou de alto nível (*high level*), é uma alternativa capaz de produzir resultados mais significativos. Porém, também é mais trabalhosa e mais propensa a erros, pois compreende atividades bem mais complexas como a codificação e a classificação de textos. Nesse método, empregam-se especialistas para codificar e classificar os textos das mensagens de acordo com um determinado conjunto de categorias, responsável por relacionar o conteúdo a uma variável de interesse. Swigger et al(2010), por exemplo, classificaram mensagens trocadas por grupos de alunos durante o desenvolvimento de projetos acadêmicos, para medir

o nível de colaboração em suas atividades e, a partir daí, determinar uma relação com o desempenho apresentado por cada grupo. Como resultado, foi possível verificar que grupos com níveis de colaboração mais altos tendem a serem os que apresentam melhores desempenhos.

2.9.2 CLASSIFICAÇÃO AUTOMÁTICA DE DOCUMENTOS

A classificação automática de documentos é um assunto amplamente estudado (LEWIS, RINGUETTE, 1994), (SCHÜTZE, HULL, PEDERSEN, 1995), (RAGAS, KOSTER, 1998), (YIMING Yang; XIN Liu, 1999), (SEBASTIANI, 2002), (SHI Yong-feng; ZHAO Yan-ping, 2004), (BAHARUDIN, LEE, KHAN, 2010). O seu rápido progresso ocorreu principalmente por propostas de técnicas de aprendizagem de máquina, representadas por algoritmos como o de Naïve Bayes, k-NN (*K-Nearest Neighbor*), SVM (*Support Vector Machines*), Árvore de Decisão (*Decision Tree*), Máxima Entropia (*Maximum Entropy*), juntamente com outras de áreas afins, como Redes Neurais, Análise Semântica Latente (*Latent Semantic Analysis* - LSA) e Correlação Nebulosa (*Fuzzy Correlation*).

Ainda assim, técnicas mais sofisticadas, como a otimização por colônia de formigas (AGHDAM, GHASEM-AGHAEE, BASIRI, 2009), a avaliação automática de ontologias para explorar conceitos semânticos (KAO, POTEET, 2007) e os algoritmos genéticos para extração de atributos significativos (BAHARUDIN, LEE, KHAN, 2010), continuam sendo propostas para aprimorar o processo de classificação, demonstrando que os trabalhos ainda estão progredindo.

Abordagens híbridas, que consideram a combinação de algoritmos, também foram propostas para a classificação automática de documentos e representam uma linha de pesquisa farta a ser explorada (GAMA, 1999), (DIETTERICH, 2000), (KUNCHEVA, 2004). Dentre os algoritmos mais empregados para essa abordagem, encontram-se o SVM, o *Naïve Bayes* e o k-NN.

A construção de uma estrutura de dados (*feature*) adequada para representar os documentos e o desenvolvimento de um classificador com alta precisão são as duas principais etapas para a classificação automática de textos (BAHARUDIN, LEE, KHAN, 2010). Basicamente, essa classificação pode ocorrer de três maneiras: não supervisionada, supervisionada e semisupervisionada.

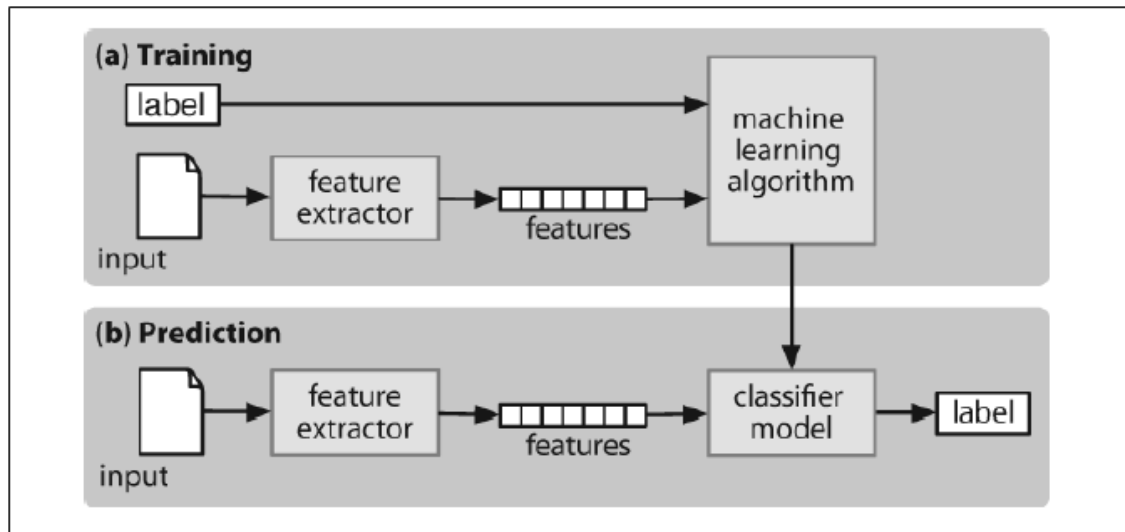


Figura 6. Classificação supervisionada. Fonte: (BIRD, KLEIN, LOPER, 2009)

A classificação supervisionada, em geral, é a mais usada. Nela, os algoritmos constroem automaticamente um classificador, aprendendo características das categorias empregadas a partir de um conjunto de treinamento com documentos já pré-classificados e, então, usa esse classificador para rotular novos documentos de acordo com as categorias aprendidas (GIORGINO, 2004).

2.9.3 ALGORITMOS DE CLASSIFICAÇÃO

Três algoritmos bastante difundidos na literatura sobre classificação de textos são: *Naïve Bayes*, *Árvore de Decisão (Decision Tree)*, *Máxima Entropia (Maximum Entropy)*.

Além de serem simples, esses algoritmos também são normalmente referenciados com desempenhos satisfatórios.

2.9.3.1 O ALGORITMO ÁRVORE DE DECISÃO

O algoritmo *Árvore de Decisão* refaz a classificação dos documentos de treinamento através da construção de uma estrutura de busca em árvore bem definida. Nessa estrutura, as folhas representam as categorias correspondentes de documentos e os ramos representam características que levam a essas categorias (Figura 7).

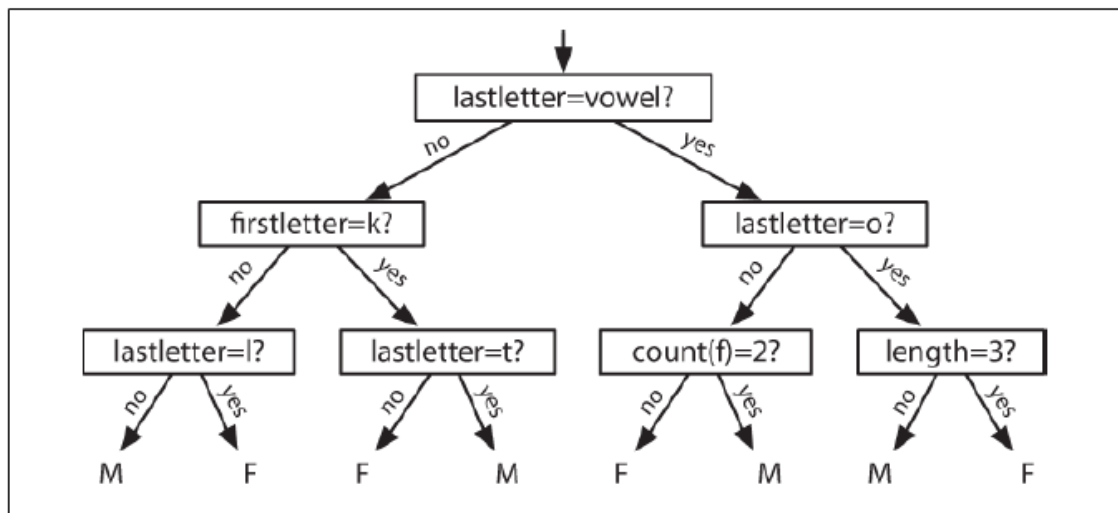


Figura 7. Diagrama representativo do algoritmo Árvore de Decisão. Fonte (BIRD, KLEIN, LOPER, 2009)

Para realizar a classificação, o algoritmo percorre a estrutura de busca a partir do nó raiz da árvore até se alcançar uma determinada folha, o que representará a categoria a ser considerada (BIRD, KLEIN, LOPER, 2009).

Esse algoritmo apresenta algumas vantagens. A principal, talvez seja a simplicidade e facilidade para interpretá-lo. Os resultados podem ser facilmente compreendidos por meio de operações simples de matemática, fornecendo uma visão consolidada da lógica usada na classificação.

Por outro lado, possui também algumas desvantagens. Seu desempenho é fortemente impactado pela quantidade de ramificações da estrutura de busca. Como a classificação textual frequentemente envolve um número grande de atributos relevantes (*features*), o número de folhas usadas para representação dos textos também é alto, o que dá origem a uma estrutura complexa e resulta em queda de desempenho. No entanto, quando há um número pequeno de atributos estruturados, como na personalização de anúncios de páginas da Internet, o desempenho é alto.

Talvez um dos grandes problemas para a implementação de uma árvore de decisão seja o *overfitting*. Com a expansão da estrutura de busca, a quantidade de dados disponível para treinar os nós mais baixos na árvore pode se tornar insuficiente, fazendo com que idiosincrasias do conjunto de treinamento sejam aprendidas ao invés de padrões linguisticamente significativos, o que resultará em classificações equivocadas.

Outro problema é a necessidade de se verificar os atributos em uma ordem específica.

Esse fato limita a exploração dos atributos ainda que sejam relativamente independentes. O algoritmo de Bayes consegue superar esta limitação, permitindo que todos os recursos sejam acessados em paralelo.

2.9.3.2 O ALGORITMO DE NAÏVE BAYES

O algoritmo de Naïve Bayes, ou Bayesiano, recebe esse nome por ter como princípio fundamental o Teorema de Bayes (WEISS et al, 2005). É, portanto, um algoritmo estatístico, que rotula um objeto de acordo com a probabilidade de esse objeto pertencer a uma das categorias consideradas.

É eficiente quando aplicado a grandes volumes de dados. E, dependendo da precisão do modelo probabilístico, consegue estimar satisfatoriamente os parâmetros necessários para a classificação a partir de conjuntos pequenos de treinamento.

Para facilitar os cálculos da classificação, admite como hipótese a independência entre os valores dos atributos, ou seja, o valor de um atributo não tem influência sobre os demais.

Essa suposição de independência traz restrições à aplicação dos classificadores Bayesianos. Ainda assim, apresentam bons resultados em diversas situações complexas do mundo real (BAHARUDIN, LEE, KHAN, 2010). Em geral, conseguem ser robusto o suficiente para suportar deficiências em suas estimativas.

Quando comparados a outros classificadores, como o SVM e Máxima Entropia, os classificadores Bayesianos apresentam desempenho relativamente mais baixo. Diversas pesquisas têm trazido novas abordagens para aperfeiçoar a classificação Bayesiana com implementações de técnicas mais efetivas (SANG-BUM KIM et al, 2002).

2.9.3.3 O ALGORITMO DE MÁXIMA ENTROPIA

O algoritmo de Máxima Entropia utiliza um modelo de classificação muito parecido com o empregado pelo classificador Bayesiano (NIGAM, LAFFERTY, McCALLUM, 1999).

A diferença está na forma de como determina os parâmetros desse modelo. Ao invés de assumir a dependência estatística das variáveis, como o classificador Bayesiano, realiza buscas para encontrar os parâmetros que maximizam o desempenho do classificador

(RATNAPARKHI, 1997). Parte do princípio que a distribuição das probabilidades de deve ser a mais uniforme possível quando nada é conhecido, ou seja, de máxima entropia.

Assim, os parâmetros recebem inicialmente valores aleatórios e, por meio de técnicas de otimização iterativas, vão sendo refinados. O refinamento garante que os parâmetros serão aproximados dos valores ótimos, mas sem determinar quando isso ocorrerá.

Por causa disso, o classificador pode levar muito tempo para aprender. Geralmente, isso acontece quando o tamanho do conjunto de treinamento, o número de atributos (*features*) e o número de categorias são grandes.

2.9.4 NECESSIDADE DE AUTOMATIZAÇÃO

A análise de conteúdo tem seus próprios desafios. Para que seja efetiva, há pelo menos dois requisitos primordiais. Em primeiro lugar, os pareceres individuais dos especialistas devem ser consistentes para que haja concordância. Sem essa garantia, não terão utilidade alguma. Adicionalmente, o resultado final deve ser alcançado em tempo suficiente para ser aplicado com oportunidade (KRIPPENDORFF, 2003).

A codificação e a classificação são atividades que, se realizadas de forma manual, normalmente consomem muito tempo, impedindo que os resultados alcançados sejam aproveitados para a mediação durante as conversações (KAO, POTEET, 2007).

Recentes avanços nas áreas de recuperação da informação e de mineração de dados têm indicado que o emprego de algoritmos de aprendizagem de máquina pode mudar esse cenário, por meio da classificação automática das mensagens. Idealmente busca-se uma mistura de métodos qualitativos e quantitativos, de modo a reforçar suas vantagens e reduzir suas limitações (BAHARUDIN, LEE, KHAN, 2010).

CAPÍTULO 3: CONTEXTUALIZAÇÃO CIRCUNSTANCIADA

3.1. A EVOLUÇÃO DA PESQUISA

A partir da caracterização do problema desta pesquisa, caminhos distintos foram percorridos com idas e vindas, proporcionando experiências que serviram para realçar a natureza complexa do processo de ensino e aprendizagem e dar um novo rumo à proposta de trabalho.

Como uma forma de auxiliar outras pesquisas, achamos importante fazer uma descrição à parte das etapas vivenciadas, com seus objetivos, atividades e resultados.

3.1.1. O MODELO BÁSICO DO EU-TU

Mediar e avaliar são tarefas fundamentais para que um professor possa desempenhar seu papel. Por intermédio delas é possível influenciar o desenrolar das ações e das reações que ocorrem no processo de ensino e aprendizagem. Quando aplicadas adequadamente, constituem um instrumento de transformação poderoso.

A avaliação mediadora proporciona autonomia e eleva o processo de ensino e aprendizagem muito além da transmissão e recepção de informações. Seus indicadores permitem acompanhar a evolução das interações e, com isso, corrigir erros e reforçar acertos, proporcionando desenvolvimento individual e coletivo.

Pensando assim, elaboramos um esquema que associa mediação (ação — orientar) com avaliação (reação — diagnóstico) para promover o processo de ensino e aprendizagem centrado em interações.

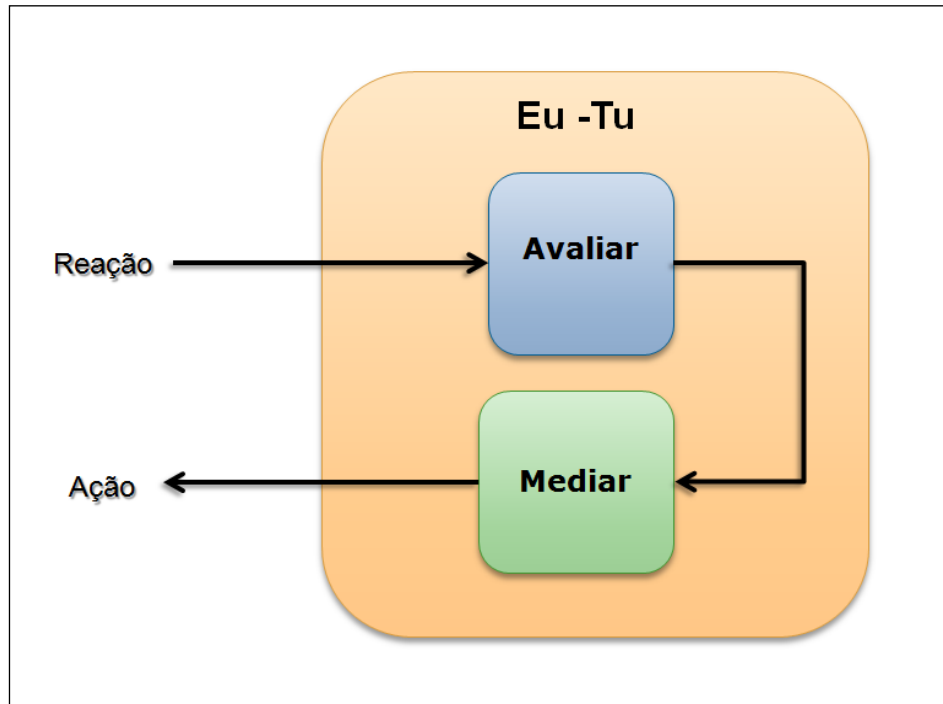


Figura 8. Avaliação Mediadora.

A partir desse esquema, construímos um modelo básico, simples e iterativo, para ser aplicado em um contexto de EAD, onde o professor proporá o desenvolvimento de tarefas de base colaborativa aos alunos.

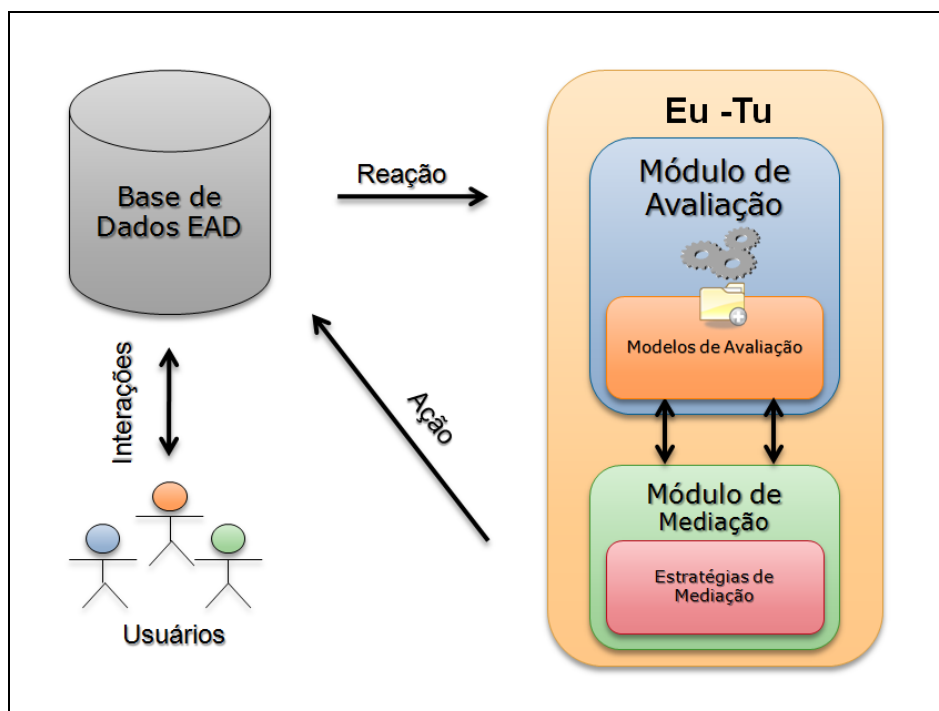


Figura 9. Modelo básico do Eu-Tu.

O Módulo de Avaliação tem por finalidade verificar a qualidade das interações escritas e gerar indicadores do progresso da aprendizagem para que seja procedida a mediação adequada oportunamente. Essa verificação pode ser feita considerando diferentes aspectos (Modelos de Avaliação), proporcionando múltiplas perspectivas para se estudar as várias dimensões do processo de ensino e aprendizagem.

O Módulo de Mediação é responsável por delinear e implementar ações (Estratégias de Mediação) que estimulem interações efetivas a partir dos indicadores gerados pelo Módulo de Avaliação.

3.1.2. O EU-TU COMO UM SISTEMA COLABORATIVO DE PERGUNTAS E RESPOSTAS

Nosso ponto de partida para aplicar o modelo básico do *Eu-Tu* foi uma realidade social em constante transformação, complexa, construída coletivamente através de interações contínuas, a partir de dúvidas e inquietações. Tínhamos como premissa que o desempenho de novos papéis poderia ser influenciado pela habilidade de questionar e responder (PEDROSA DE JESUS, 2009; MOREIRA, 1999). Isso significa que não bastaria apenas fazer perguntas por si só. Seria necessário que elas fossem feitas em uma sequência lógica (RICHETTI, SHEERIN, 1999). Também consideramos que, ao questionar, as pessoas expressam todo o seu “eu” (ROGERS, 1960a, 1960b), inclusive suas preferências de aprendizagem e de ensino. Parodiando um dito popular, esta premissa seria análoga à: “*diga-me como perguntas (ou respondes) e eu te direi como aprendes (ensinas)*”.

Assumimos, então, que o emprego de um sistema automático de perguntas e respostas poderia ser uma ação pedagógica efetiva para esse contexto. Seguindo o modelo básico elaborado, propomos usar recomendações de perguntas e respostas relacionadas ao tema em questão (ações) como Estratégia de Mediação e o grau de satisfação dos usuários (reações) como Modelo de Avaliação.

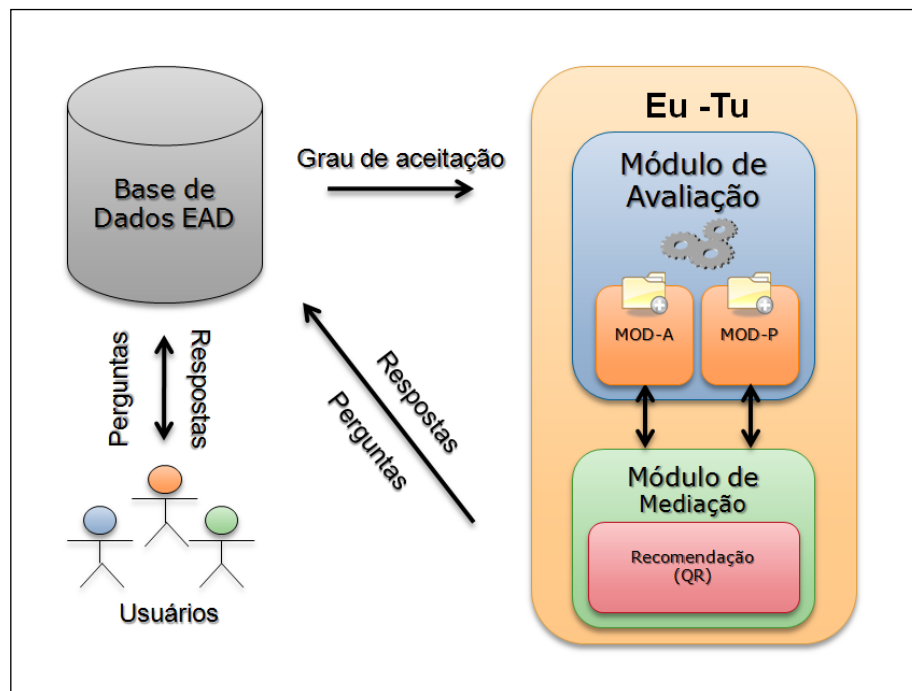


Figura 10. Eu-Tu instanciado como um Sistema de Perguntas e Respostas.

Tratava-se de explorar uma forma indireta de mediação, que se apoia na habilidade de uns questionarem e de outros responderem. Esses papéis poderiam ser trocados dependendo do assunto questionado, em uma relação dialética análoga àquela proposta por Hegel (MENESES, 2003) para explicar a evolução do espírito humano.

O diagrama da Figura 11 ilustra o funcionamento do sistema automático de perguntas e respostas proposto. Conforme esse diagrama, Alunos A1, A2,..., An, de determinado perfil fazem perguntas que são respondidas por professores P1, P2,..., Pm, também com um dado perfil, em um fórum virtual (FAQR), para discutir e negociar a compreensão de um assunto de interesse comum.

Com a evolução das interações, é criado um conjunto base (BQR) de perguntas e respostas que, de um lado, serve para recomendar, por meio de um processo de filtragem, um subconjunto (SQR) de perguntas e respostas relacionadas com o assunto em questão, substituindo a mediação direta. O professor somente auxiliará quando as respostas recomendadas não forem consideradas satisfatórias.

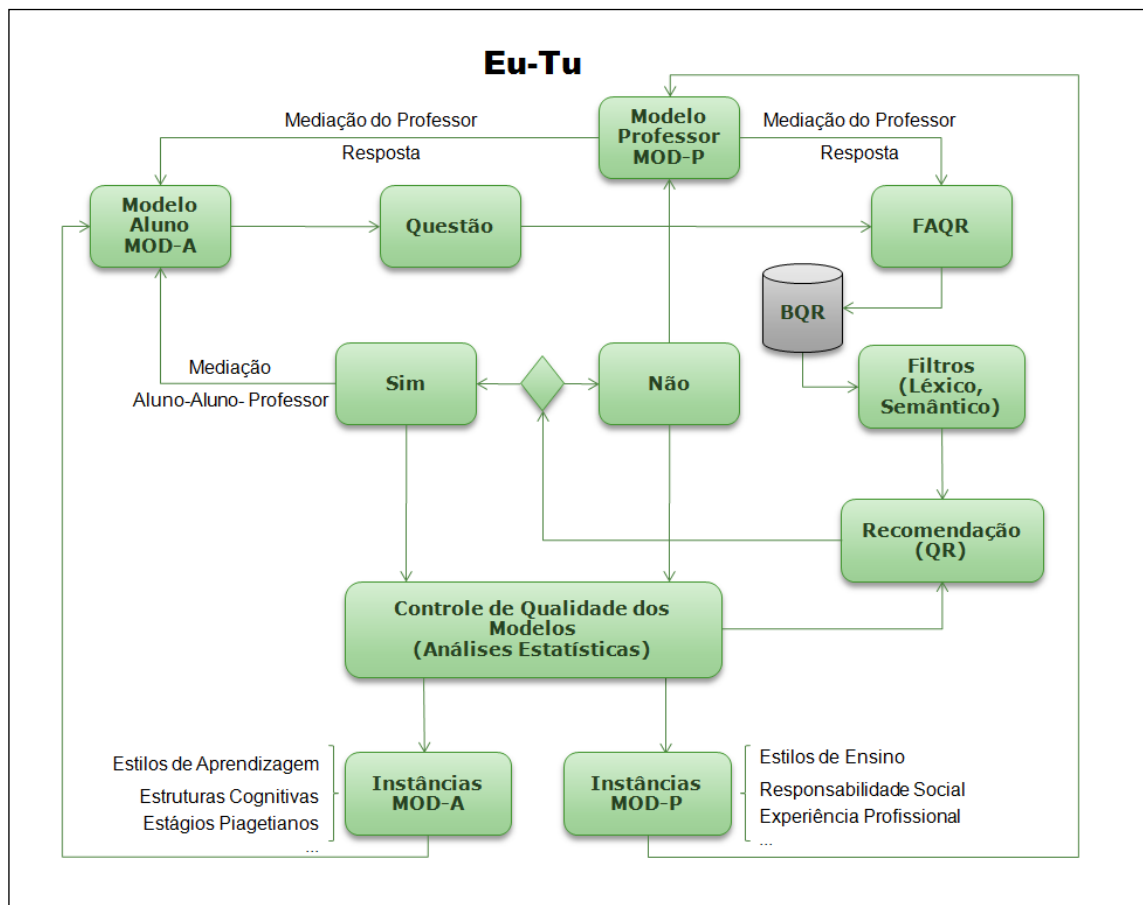


Figura 11. Diagrama do Sistema de Perguntas e Respostas

De outro lado, mas não de forma independente da anterior, o conjunto base (BQR) também serve para identificar como as características individuais influenciam e são influenciadas. A partir das interações de grupos de usuários reais com o sistema, modelos representativos dos alunos (MOD-A) e dos professores (MOD-P) são construídos e, paulatinamente, aperfeiçoados através de análises estatísticas diversas. Com isso, é dado início a um processo recursivo de aprimoramento, que permitirá usar esses modelos para a mediação, ao invés de usuários reais, sendo que estes continuarão a atuar como provedores de novas informações.

Duas das principais exigências do sistema proposto eram colaboração e adaptação. Logo, professores e alunos precisavam ter conhecimento pleno de suas características, de forma que escolhessem as melhores estratégias a serem aplicadas em busca de interações efetivas. Por exemplo, não seria interessante sempre privilegiar uma mesma característica, pois a capacidade de interagir é aumentada com o desenvolvimento de novas habilidades. Aqueles que não tivessem desenvolvido habilidades adequadas para se adaptarem teriam mais dificuldades para interagir. Diante disso, decidimos incluir os estilos de aprendizagem e de

ensino no Modelo de Avaliação, por considerarmos que a eficiência para lidar com as diversas formas de como as informações são apresentadas é aumentada à medida que novos estilos são desenvolvidos (FELDER, SPURLIN, 2005). Segundo Cavellucci (2002), os estilos de ensinar são instrumentos capazes de identificar e modelar as preferências dominantes na forma como as pessoas recebem e processam informações para aprender e ensinar.

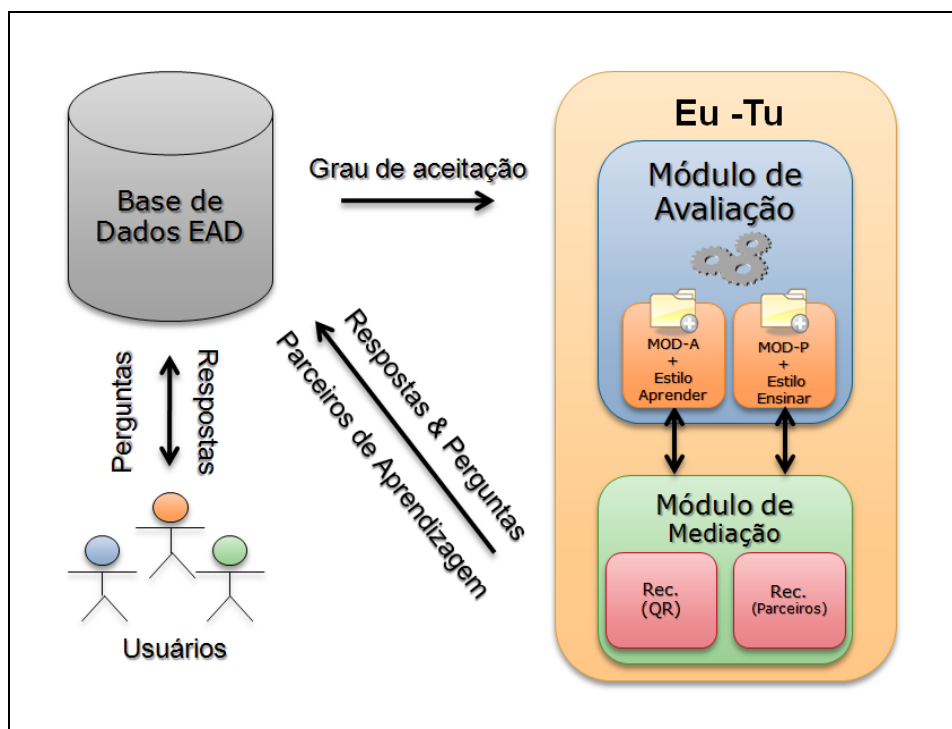


Figura 12. Eu-Tu instanciado como um Sistema de Perguntas e Respostas, com Combinação Social.

Adicionamos também a Combinação Social como uma nova Estratégia de Mediação. Nossa meta era criar uma ferramenta capaz de identificar os estilos de alunos e professores e, por meio da recomendação de parceiros que estimulassem o desenvolvimento de novas habilidades, minimizar os efeitos causados pelas diferenças entre esses estilos.

Para instanciar esse modelo por um sistema informatizado de perguntas e respostas, focamos na construção de um mecanismo capaz de recuperar, a partir do conjunto base de perguntas e respostas (BQR), o subconjunto de respostas que mais se relacionavam com o assunto em questão (SQR).

Utilizamos o *framework Apache Lucene*¹², que é uma biblioteca Java, gratuita e *open-source*, largamente usada para recuperação de informação. Com o *Lucene*, foi possível

¹² Site oficial: <http://lucene.apache.org/>.

acrescentar os recursos de pesquisa de texto necessários ao nosso sistema.

3.1.3. ESTUDOS COM ESTILOS DE APRENDIZAGEM

Apesar de haver similaridades e complementações entre os diversos tipos de estilos de aprendizagem, eles definem "diferenças diferentemente". A multiplicidade de definições e modelos propostos acaba gerando dificuldades para se trabalhar com os estilos de aprendizagem (DUNN et al, 1981; GUILD, GARGER, 1985; DESMEDT, VALCKE, 2004). Estudamos então três dos principais modelos de estilo de aprendizagem: o *Myers-Briggs Type Indicator 1* - MBTI-1 (MYERS, 1998), o *Index of Learning Styles* - ILS (FELDER, SILVERMAN, 1988) e o *Learning Style Inventory* - LSI (KOLB, 1984), para entender com maior clareza como eles caracterizam os indivíduos e verificar se existiram restrições ao empregá-los para flexibilizar relações sociais estabelecidas no processo educacional.

Observamos que os três modelos estudados realmente possuem similaridades quanto às categorizações dos estilos de aprendizagem. Basicamente, relacionam os estilos a quatro tipos de categorias: personalidade, processamento da informação, interações sociais e métodos instrucionais. A dimensão Sensitivo–Intuitivo, do ILS, é a mesma no MBTI e apresenta equivalência com a dimensão Experiências Concretas–Conceituação Abstrata (EC–CA), do LSI. A dimensão Ativo – Reflexivo, do ILS, se correlaciona com as dimensões Extroversão–Introversão, do MBTI, e Reflexiva–Experimentação Ativa (OR–EA), do LSI. Concluímos que o ILS além de ser o modelo mais empregado nos sistemas adaptativos com base em estilos de aprendizagem, leva em consideração as interações sociais como um critério para a classificação de seus estilos (RUIZ et al, 2008; PAPANIKOLAOU, GRIGORIADOU, 2004).

Chamou-nos a atenção o fato de nenhum dos três modelos estudados considerar o estágio de desenvolvimento dos aprendizes, nem aspectos conativos, relativos aos desejos e às intenções, nem afetivos. Percebemos que muito do que se tem pesquisado a respeito das diferenças na aprendizagem permanece amplamente focado nos mecanismos, intrínsecos ou extrínsecos, de processamento da informação e de construção do conhecimento. Com isso, identificamos que um fator crítico para os sistemas adaptativos de aprendizagem é a determinação de outras fontes dominantes, de alto nível, para as diferenças individuais de aprendizagem (MARTINEZ, BUNDERSON, 2008). Por exemplo, isso significa compreender como intenções, desejos, sentimentos e emoções influenciam os processos cognitivos. Por

consequência, seria então necessário estender nossa pesquisa para também incluir influências conativas, afetivas e sociais.

Analizamos ainda alguns trabalhos para identificar como os estilos de aprendizagem podem ser empregados na adaptação de sistemas educacionais.

Vimos que existem basicamente 03 (três) propostas (PAPANIKOLAOU, GRIGORIADOU, 2004):

- Adaptação (tipo e sequência) de conteúdo instrucional: os estilos de aprendizagem são utilizados para identificar as preferências e as estratégias empregadas pelos aprendizes a fim de que lhe sejam apresentados conteúdos adequados as suas características, como, por exemplo, o tipo de mídia preferida (CANDOTTI et al., 2006; BARCELLOS et al., 2007);
- Navegação adaptativa: os estilos de aprendizagem servem para identificar as preferências dos aprendizes quanto à estrutura dos recursos instrucionais a serem utilizados, de modo a orientá-los na organização de suas rotas de aprendizagem, como, por exemplo, indicação de atividades individuais ou em grupos (HUMMEL et al., 2006);
- Colaboração adaptativa: os estilos de aprendizagem são empregados para formar grupos de alunos em contextos de aprendizagem colaborativa, uma vez que tais características são consideradas capazes de influenciar as interações sociais. Ex.: formação de grupos com a maior diversidade possível de estilos de aprendizagem (ALFONSECA et al., 2006; PAREDES, RODRIGUEZ, 2006).

Concluímos que nossa pesquisa estava mais relacionada com as tecnologias de colaboração adaptativa e de navegação adaptativa.

Observamos também que os estilos de aprendizagem são ainda divulgados e empregados de forma pouco apropriada entre os educadores (FUTER, 2007). Muitos professores ensinam e orientam o estudo de acordo com seus próprios estilos, sem conhecer os estilos de aprendizagem de seus alunos. Para Santos, Bariani, e Cerqueira (2000), por exemplo, melhores resultados seriam alcançados se, nos cursos de formação de professores, houvesse pelo menos uma introdução ao estudo da conceituação dos estilos de ensinar e de aprender.

Desses estudos, fomos capazes de identificar alguns pontos críticos que influenciam o

desenvolvimento de sistemas adaptativos educacionais baseados em estilos de aprendizagem. Em suma, são eles:

- a seleção de um modelo dentre os diversos propostos para serem empregados na representação dos estilos de aprendizagem;
- o emprego de tecnologias adaptativas eficientes capazes de lidar com diversos estilos de aprendizagem;
- a identificação automática dos estilos de aprendizagem dos usuários do sistema (alunos).

A partir daí, demos início a um estudo exploratório para verificar se os resultados alcançados por outros trabalhos de pesquisa (GRAF, KINSHUNK, 2009; POPESCU, BADICA, 2009; SAEED, YANG, SINNAPPAN, 2009), que tratam da identificação automática dos estilos de aprendizagem, poderiam também ser reproduzidos utilizando os fóruns eletrônicos de discussões assíncrona de um ambiente virtual de aprendizagem (AVA).

Esse estudo exploratório envolveu 14 (quatorze) alunos do curso PGTIAE¹³, durante as aulas da disciplina "Avaliação". O AVA utilizado pela disciplina foi a Plataforma Interativa para a Internet – *Pii* (ELIA; SAMPAIO, 2001), que tinha as funcionalidades requeridas pela pesquisa: um questionário *on line*, com base no ILS, para identificar os estilos de aprendizagem de todos os alunos da turma; e um fórum virtual denominado Arena. Na Figura 13, apresentamos os resultados obtidos pela aplicação do questionário nas primeiras aulas.

O estudo consistiu em propor aos alunos da disciplina que discorressem, de acordo com o seus próprios conhecimentos, sobre o tema "o que é e para que serve uma avaliação". Tal proposta tinha por finalidade medir o nível de percepção dos alunos em relação ao assunto que seria estudado. A partir dos estilos de aprendizagem identificados através do ILS e pela medida do nível de percepção dos alunos, a turma foi dividida em grupos de estudos, com 5 (cinco) alunos cada.

¹³ Curso de Pós-Graduação em Tecnologia da Informação Aplicadas à Educação do Núcleo de Computação e Eletrônica da Universidade Federal do Rio de Janeiro (NCE/UFRJ). É uma pós-graduação *Lato-Sensu* voltada para especialização em tecnologias da informação aplicadas à educação/treinamento, tanto nos aspectos tecnológicos, quanto pedagógicos.

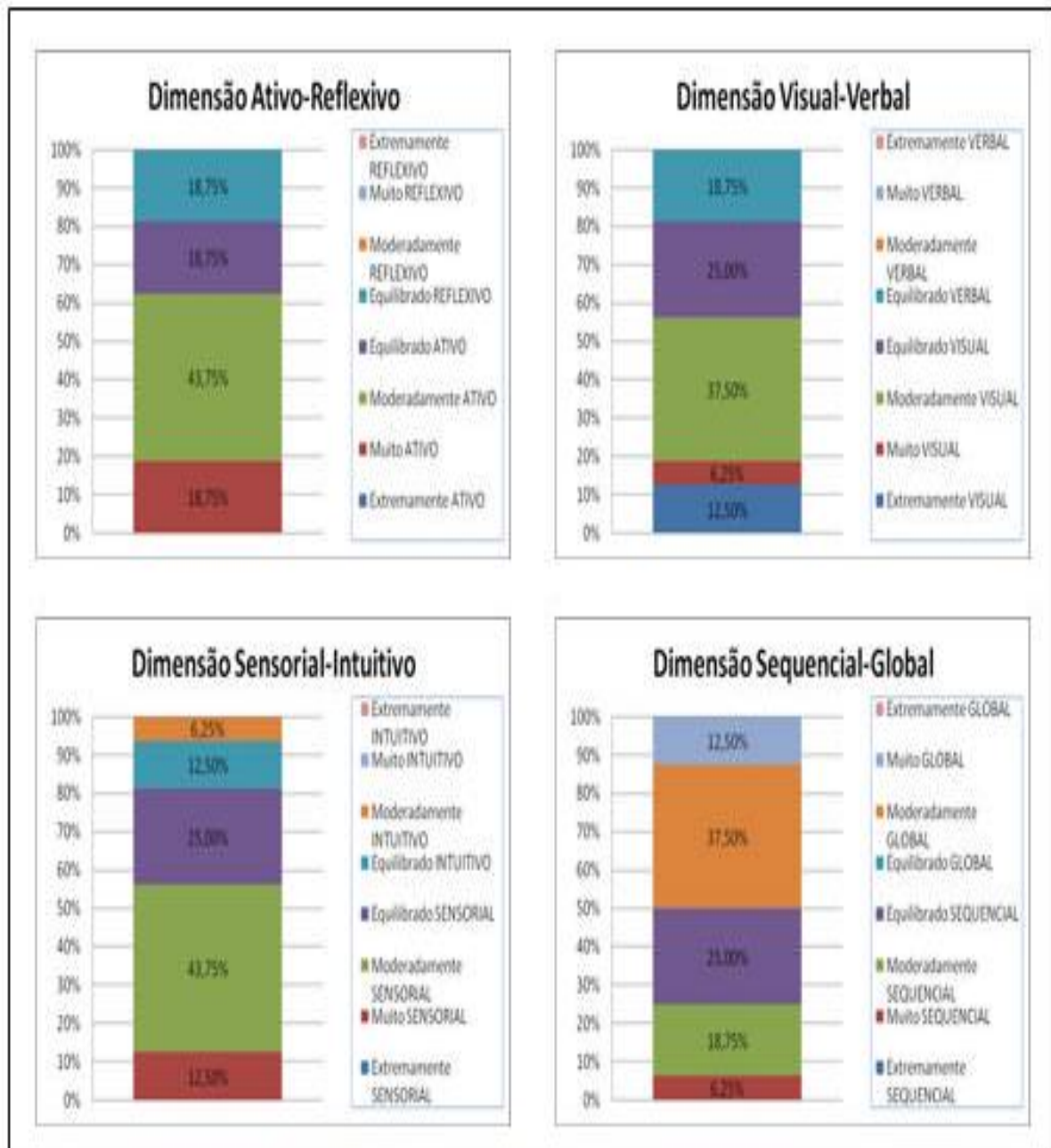


Figura 13. Distribuição dos Estilos de Aprendizagem dos alunos do PGTIAE

Feita a divisão, observamos a participação dos alunos nas salas de discussões (*Arenas*) criadas para cada grupo debater sobre o processo de avaliação na educação. Medimos a duração e o número de acessos a cada *Arena* e a quantidade total das mensagens postadas pelos alunos durante o primeiro mês do curso, com o intuito de identificar a existência de um padrão característico relacionado às dimensões ATIVO-REFLEXIVO e VISUAL-VERBAL, do modelo ILS. Nossos dados não nos permitiriam ser conclusivos em relação à existência desses padrões, mas indicaram que usuários de estilos diferentes interagem diferentemente.

Com base nesses resultados, formamos grupos homogêneos no que diz respeito a sua classificação ILS com o objetivo de buscar evidências a respeito de como aprendizes com diferentes estilos de aprendizagem utilizam um sistema de perguntas e respostas para selecionar recursos considerados mais apropriados aos seus estilos. Tínhamos como hipótese a possibilidade de classificar as perguntas e as respostas de acordo com os estilos de aprendizagem e de ensino, respectivamente.

Propusemos aos alunos o desenvolvimento de uma tarefa de base colaborativa, tendo como meta formar uma coleção de questões sobre avaliação educacional, formuladas e respondidas pelos alunos, com a revisão do professor.

Cada aluno de um dado grupo propunha 3 (três) perguntas aos pares do seu grupo, sendo que o conjunto formado era então discutido criticamente em um fórum virtual (Arena da Pii) com o intuito de selecionar e incluir as 2 (duas) perguntas consideradas mais relevantes para serem disponibilizadas no *Eu-Tu*.

Com esse conjunto de perguntas, pretendíamos realizar dois estudos específicos.

- O primeiro, para verificar a existência de uma relação entre os estilos de aprendizagem e de ensinar e os tipos de perguntas e de respostas formuladas, como por exemplo, se alunos sensoriais tendem a fazer perguntas convergentes, cujas respostas envolvem abordagens aprendidas anteriormente.
- O segundo estudo seria uma análise multivariada de classificação hierárquica (*Cluster Analysis*) para agrupar os alunos segundo as suas comunicações. Os grupos construídos a base dos estilos de aprendizagem exerceriam o papel de aprendizes formulando questões (dúvidas), que seriam respondidas pelos grupos construídos a base dos estilos de ensino (tutores). Cada grupo-tutor escolheria as questões preferidas para responder dentre as da coleção produzida pelos grupos-aprendizes.

As análises referentes ao primeiro estudo foram feitas por um grupo de alunos voluntários como trabalho final da disciplina de avaliação, sob orientação do autor principal desta pesquisa. Entretanto, os resultados obtidos não foram animadores¹⁴.

Além disso, ao apresentarmos nossa proposta de pesquisa no 1º Seminário de Acompanhamento de Trabalhos de Dissertação do PPGI, foram feitas observações que nos

¹⁴ Conforme pode ser visto em:
http://146.164.3.41/Pii2009/Projeto190/Debyte/DB_1376_7_Relatorio%20final_AVAL.doc

fizeram ver a importância de estender nosso modelo inicial, restrito até então apenas à habilidade de questionar e responder, para um centrado em interações baseadas na discussão escrita.

3.1.4. MUDANÇA DE RUMO

Uma vez que a aprendizagem ocorre por ações distintas, era uma simplificação muito forte considerar apenas aquelas que ocorrem por meio de perguntas e respostas. Passamos, então, a enxergar o *Eu-Tu* não mais como um sistema colaborativo de perguntas e respostas apenas, e sim como um sistema para análise de interações em fóruns de discussões eletrônicos, capaz de produzir indicadores para a mediação. Tecnicamente, isso representou uma mudança de abordagem. Com a classificação de mensagens associada a níveis de aprendizagem, passamos de "recuperação de informação" para "extração de informação" (MOLDOVAN, SURDEANU, 2002).

Vimos que vários modelos disponíveis na literatura para análise de interações apresentam perspectivas bem diferentes (SOLLER, LESGOLD, 2000). Apesar dessa diversidade, apontam para um consenso de que tanto o que é feito (VANLEHN et al, 1998) quanto quem participa das interações exercem influências sobre os resultados alcançados (PAREDES, RODRIGUEZ 2006; ALFONSECA, 2006; ADÁN COELLO et al, 2008; MARTINEZ, BUNDERSON, 2008; GRAF, KINSHUNK, 2009; POPESCU, BADICA, 2009; SAEED, YANG, SINNAPPAN, 2009).

Assim, para que as discussões em fóruns eletrônicos sejam conduzidas de forma que alavanquem o desenvolvimento do pensamento crítico, é preciso levar em consideração o contexto dessas discussões, o que inclui suas atividades, os estilos de seus participantes e os papéis desempenhados por eles (LOPES, 2007).

Percebemos um pouco disso nos estudos que realizamos com estilos de aprendizagem. Verificamos que existem preferências e habilidades específicas que nos fazem interagir de maneiras diferentes no processo de ensino e aprendizagem.

O ideal é que, além de argumentar, os aprendizes também desenvolvam outras habilidades, que lhes dê condições de suportar as diferenças e evitar choques que prejudiquem as interações. Para isso, a tecnologia mostra-se como uma aliada importantíssima. Ela pode fornecer ferramentas computacionais para automatizar a análise das interações e gerar indicadores que revelem a qualidade dos produtos do processo de ensino e aprendizagem,

permitindo acompanhá-lo de perto. Esses indicadores atuarão como instrumentos de auto-avaliação capazes de explicitar os resultados das ações realizadas, planejar novas atitudes e controlar o ritmo de trabalho.

Assim, chegamos à proposta atual deste trabalho de pesquisa que é explorar a troca de mensagens em um fórum eletrônico de discussões como uma forma indireta de promover o processo de ensino e aprendizado centrado em interações. A base para isso é analisar o conteúdo das mensagens, a fim de agrupar as interações segundo indicadores que identificam e avaliam a construção social do conhecimento. No Capítulo 4, fazemos uma descrição detalhada do modelo aplicado para a construção do *Eu-Tu*.

CAPÍTULO 4: A IMPLEMENTAÇÃO DO MÓDULO DE AVALIAÇÃO

Neste capítulo, apresentamos as etapas percorridas para desenvolver e implementar um protótipo para o Módulo de Avaliação do *Eu-Tu*.

Começamos pela descrição do Modelo de Avaliação escolhido para o *Eu-Tu* (seções 4.1 e 4.2) e apresentamos o diferencial de nossa implementação (seção 4.3).

A partir da seção 4.4, detalhamos as duas etapas do desenvolvimento do Módulo de Avaliação. A primeira foi a determinação de um algoritmo de classificação (seção 4.4.1). Realizamos um estudo piloto que permitiu evidenciar resultados de trabalhos apresentados na literatura sobre classificação automática de textos e que serviram para orientar a seleção do algoritmo. Esse estudo é descrito em detalhes, com suas finalidades específicas, procedimentos, resultados e conclusões, com o intuito de explicar porque selecionamos o classificador Bayesiano. A segunda etapa foi a construção do classificador automático (seção 4.4.2). Para isso, restava determinar a melhor estratégia de classificação. Estudamos então a influência do tamanho do conjunto de treinamento e do número de categorias sobre o desempenho do classificador Bayesiano (seções 4.4.2.1 e 4.4.2.2). Esse novo estudo piloto nos levou finalmente a concluir que seria melhor empregar um conjunto (*ensemble*) de classificadores binários como estratégia para a implementação do Módulo de Avaliação do *Eu-Tu*.

4.1. PROMOVENDO O PROCESSO DE ENSINO E APRENDIZADO CENTRADO EM INTERAÇÕES ATRAVÉS DA TROCA DE MENSAGENS EM FÓRUMS ELETRÔNICOS DE DISCUSSÕES

As discussões realizadas em fóruns eletrônicos podem ser analisadas a fim de avaliar o processo de aprendizagem e contribuir para a construção colaborativa do conhecimento (LUI et al 2004; SWIGGER et al, 2010). Entretanto, é preciso que essa análise seja mais que uma medida quantitativa do conteúdo adquirido em um certo momento. Ela deve ocorrer sistematicamente e considerar o desempenho do aluno nas diferentes circunstâncias que ocorrerem durante o processo de ensino e aprendizagem. Por essa razão, propomos empregar

o modelo básico do *Eu-Tu* para criar uma ferramenta que explora a troca de mensagens em fóruns eletrônicos de discussões para promover o processo de ensino e aprendizagem centrado em interações.

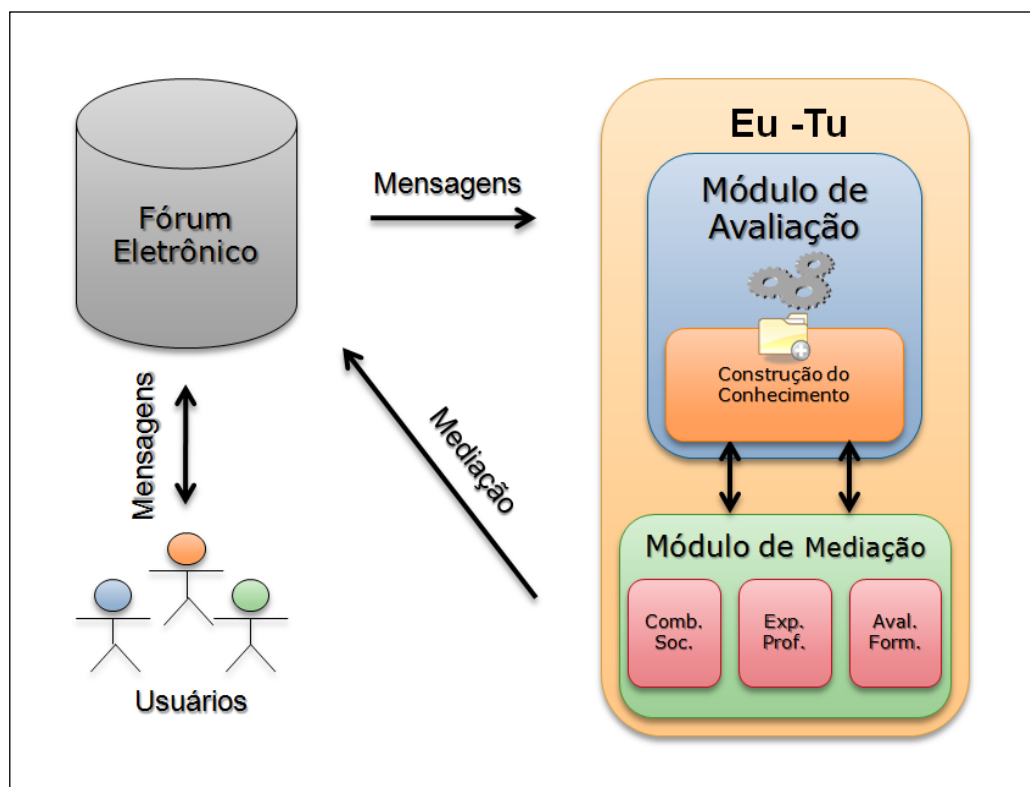


Figura 14. Uma instância do modelo básico do *Eu-Tu* aplicada em Fóruns Eletrônicos de Discussões.

Nesse novo esquema, adotamos a construção do conhecimento como indicador de aprendizagem e, por isso, escolhemos a taxonomia desenvolvida por Lopes (2007) para o Modelo de Avaliação. Dessa forma, as interações são agrupadas conforme as fases da evolução das discussões, o que permite identificar a construção social do conhecimento e fazer considerações sobre a dinâmica do processo ensino-aprendizagem.

Admitimos também um conjunto de estratégias no Módulo de Mediação, que inclui a combinação social e a experiência do professor, para tratar das diferentes características dos aprendizes e desenvolver habilidades necessárias à inovação.

Como o objetivo específico deste trabalho prioriza a geração automática de indicadores de aprendizagem, tratamos apenas do desenvolvimento do Módulo de Avaliação. Nas próximas seções, detalharemos como se deu esse desenvolvimento através da construção de um classificador automático de textos para analisar o conteúdo de mensagens de fóruns

eletrônicos de discussões.

4.2. O MODELO DE AVALIAÇÃO ESCOLHIDO PARA O EU-TU

Dentre alguns trabalhos que pesquisamos a respeito da avaliação do processo de ensino-aprendizagem mediante a análise do conteúdo de interações em fóruns de discussões, destacamos o trabalho de Lopes (2007). Chamou-nos a atenção a forma como a autora encara a avaliação. Para ela, a avaliação compreende *“parte do processo de aprendizagem, vislumbrando-se incorporar sua riqueza no sentido de proporcionar desenvolvimento por meio da construção de conhecimento, bem como melhorias conforme as necessidades do aluno individual e coletivamente”*. Assim, Lopes (2007) construiu uma taxonomia capaz de classificar as interações de fóruns eletrônicos de discussões em categorias semânticas que contemplam indicadores de aprendizagem.

Para desenvolver seu trabalho, Lopes (2007) considerou o método proposto por Gunawardena (1997), o qual pressupõe que as discussões podem passar por uma série de estágios, capazes de explicar um padrão de construção de conhecimento manifestado durante o debate (ANEXO 1).

Lopes (2007) afirma que esse método foi escolhido porque consta de importantes características para a compreensão do processo de aprendizagem em atividades colaborativas em EAD:

- considera as interações como veículo de co-construção de conhecimento;
- estabelece a existência de um padrão de construção de conhecimento a partir de um fórum;
- considera contexto de aprendizagem com base em teorias sócio-interacionistas;
- apresenta um esquema relativamente simples e direto;
- é adaptável para o estabelecimento de uma escala de medida em um contexto de ensino aprendizagem.

Após uma série de experimentos, Lopes (2007) selecionou e validou¹⁵ inícios de sentenças representativos para classificar as interações de acordo com estágios propostos por

¹⁵ A taxonomia foi convalidada para uma atividade didática que corresponde à leitura e compreensão de um artigo científico por meio de um AVA, envolvendo uma amostra de profissionais com ampla experiência em EAD. Outros escopos requererão provavelmente outras taxonomias.

Gunawardena (1997). Esses inícios de sentenças foram então agrupados em categorias¹⁶ semânticas e fases de aprendizagem, dando origem a taxonomia apresentada no Anexo 2.

Com essa taxonomia, Lopes (2007) conseguiu analisar, quantitativamente e qualitativamente, os arquivos dos registros das interações, armazenados tanto por tempo quanto por debatedor. As mensagens iniciadas com o texto de referência foram classificadas na taxonomia das categorias e o nível de aprendizagem, identificado. Dessa forma, foram obtidos dados que possibilitaram validar a coerência da dinâmica das discussões com as fases de aprendizagem propostas por Gunawardena (1997) e construir um fórum categorizado para a avaliação da aprendizagem.

4.3. DIFERENCIAL E OPORTUNIDADES DE MELHORIA

No fórum categorizado apresentado por Lopes (2007), toda comunicação entre professor e alunos acontece via fórum eletrônico de discussões. Antes de escrever sua mensagem, cada usuário deve selecionar, em um menu específico, o texto inicial a ser empregado e que representará o significado do que deseja expressar. Ao fazer isso, a mensagem é indiretamente pré-classificada, uma vez que o texto escolhido está relacionado à taxonomia de categorias com significados semânticos para as discussões. Outra forma também de acompanhar a evolução da discussão e entender a dinâmica do processo de ensino-aprendizagem é através da pós-classificação. Nela, as mensagens são armazenadas para que sejam analisadas e classificadas futuramente, em geral, por especialistas.

Apesar de, em ambos os casos, ser possível associar os registros das interações às fases de aprendizagem, o emprego desses dois tipos de classificação não ocorre indistintamente. Existem aspectos que podem favorecer mais a um tipo e desfavorecer o outro. Por exemplo, a pré-classificação pode proporcionar mediação quase que de imediato, entretanto não é muito precisa. Em experimentos iniciais, encontramos várias discrepâncias entre o significado do conteúdo das mensagens e a classificação que receberam de seus autores. Notamos também que as classificações mais precisas foram as geradas por

¹⁶ As categorias e subcategorias escolhidas originaram-se da seleção de situações ocorridas em uma atividade colaborativa que pressupõe discussão, troca de informações, possibilidade de conflito e negociação para resolver o problema apresentado e da identificação, designando um nome para as situações semelhantes. Lopes (2007) reconhece que as situações possíveis de ocorrerem em um debate não foram esgotadas, mas afirma que as categorias escolhidas foram as que emergiram das situações estudadas e se mostraram semelhantes àquelas verificadas nos autores que utilizam classificação de sentenças para organizar e atribuir significados às interações de um discurso.

especialistas, como professores ou mediadores. Parece que há obstáculos que impedem os autores de escrever o que realmente intencionavam. Tal fato pode estar relacionado a problemas de interpretação, atenção baixa ou mesmo ausência de motivação para envolvimento com as tarefas propostas e conduz a resultados equivocados e comprometedores, principalmente quando se trata de avaliação. Por outro lado, a pós-classificação, se realizada de forma manual, demanda muito tempo. As mensagens precisam ser armazenadas em *Logs* para serem analisadas posteriormente às discussões. Esse desperdício retarda a mediação, impedindo-a de ser feita apropriadamente.

Essa dicotomia "tempo-precisão" nos pareceu estar ainda mais intensificada em nosso trabalho e ter grande relevância para termos bem sucedidos, uma vez que planejamos promover o processo de ensino e aprendizagem por meio de uma avaliação mediadora proporcionada pela análise de interações escritas.

Foi preciso encontrar uma maneira de aliar os benefícios e reduzir os prejuízos oriundos da pré-classificação e pós-classificação. A classificação automática se mostrou um bom caminho a seguir e, por isso, a incluímos no modelo proposto. Entendemos que um classificador automático é um importante componente para o Módulo de Avaliação do *Eu-Tu*, capaz de proporcionar índices de avaliação adequados e, simultaneamente, reduzir, ou mesmo eliminar, o desperdício de tempo para gerar esses índices. Com isso, o professor terá condições para mediar oportunamente, enquanto as interações estiverem ocorrendo.

4.4. DESENVOLVIMENTO DO MÓDULO DE AVALIAÇÃO

O desenvolvimento do Módulo de Avaliação se deu em duas etapas, que envolveram diferentes atividades. A primeira teve como objetivo determinar um algoritmo de classificação apropriado para o escopo da pesquisa. A segunda etapa foi a construção do engenho de classificação, considerando a maximização de seu desempenho, no que se refere ao tempo de processamento e à precisão.

4.4.1. ETAPA I: DETERMINAÇÃO DO ALGORITMO DE CLASSIFICAÇÃO

Iniciamos essa etapa realizando um estudo, não exaustivo, de trabalhos sobre classificação de documentos e mineração de dados textuais, para levantar os principais tipos de classificadores que poderiam ser empregados. Nos concentramos principalmente na

representação de textos e mecanismos de aprendizagem de máquina, destacando as técnicas e metodologias mais utilizadas e, ao mesmo tempo, identificando alguns de seus desafios e limitações. Esse estudo também serviu para a fundamentação teórica desta pesquisa. Como resultado, vimos que dentre os algoritmos de classificação mais citados na literatura estão: *Naïve Bayes*, *Árvore de Decisão (Decision Tree)*, *Máxima Entropia (Maximum Entropy)*.

Em seguida, realizamos dois experimentos empregando técnicas de Processamento de Linguagem Natural e um conjunto de ferramentas open source, escritas em *Python*, chamado de *Natural Language Toolkit – NLTK* (NLTK, 2010), para auxiliar a escolha de qual algoritmo usar dentre os que havíamos estudado.

4.4.1.1. EXPERIMENTO 1: ANÁLISE DE DESEMPENHO

No primeiro experimento, analisamos empiricamente o desempenho de três algoritmos para classificadores binários, Bayesiano, *Árvore de Decisão* e de *Máxima Entropia*, em relação à precisão e ao tempo de processamento. Nosso interesse foi verificar a capacidade desses algoritmos para prever corretamente a classe de uma instância. Para isso, usamos recursos do próprio NLTK, conforme mostra a Figura 15.

```
def aval_classifier(classifier, testfeats):
    refsets = collections.defaultdict(set)
    testsets = collections.defaultdict(set)

    for i, (feats, label) in enumerate(testfeats):
        refsets[label].add(i)
        observed = classifier.classify(feats)
        testsets[observed].add(i)

    print 'precisão:', nltk.classify.util.accuracy(classifier, testfeats)
    print 'pos precision', nltk.metrics.precision(refsets['pos'], testsets['pos'])
    print 'pos recall:', nltk.metrics.recall(refsets['pos'], testsets['pos'])
    print 'neg precision:', nltk.metrics.precision(refsets['neg'], testsets['neg'])
    print 'neg recall:', nltk.metrics.recall(refsets['neg'], testsets['neg'])
    classifier.show_most_informative_features()
```

Figura 15. Código para medir a precisão dos classificadores

Com esse código, os erros cometidos pelos classificadores foram automaticamente agrupados de acordo com uma matriz de confusão binária (COLAS, 2009), conforme ilustrado na Tabela 1.

Tabela 1. Matriz confusão binária

		CLASSES PREVISTAS	
		A	B
CLASSES CONHECIDAS	A	tp_A	e_{AB}
	B	e_{BA}	tp_B

Os símbolos tp_A e tp_B representam respectivamente os números de classificações corretas (*True Positives*) para as classes A e B. Os erros de classificações (*False Positives*) são representados por e_{AB} e e_{BA} .

A partir dessa matriz, as precisões das classificações foram calculadas empregando-se a fórmula abaixo:

$$Precisão_A = \frac{tp_A}{tp_A + e_{BA}}$$

Figura 16. Fórmula da precisão para uma classe "A" hipotética.

O tempo de processamento compreendeu o período para a realização das atividades de treinamento e de teste, medido em segundos.

Para o experimento, utilizamos um *notebook* Intel® Core™ 2 Duo CPU P7550@2.26 GHz, com 8 GB de memória RAM e sistema operacional Windows 7®, Home Premium, 64 bit, com *Service Pack 1*.

Usando o NLTK como ferramenta, construímos três classificadores binários; um para cada tipo escolhido. O código da Figura 17 demonstra como criamos o classificador Bayesiano.

```
def train_classifier(trainfeats):
    classifier = NaiveBayesClassifier.train(trainfeats)
    return classifier
```

Figura 17. Código para criação de um classificador Bayesiano binário

Também aproveitamos o *corpus* "Movie Reviews", da base nativa do NLTK, como conjunto de dados para o experimento. Esse *corpus* possui 2000 arquivos de textos, igualmente distribuídos entre duas classes, "positivo" e "negativo", com comentários a respeito de filmes.

Formamos os conjuntos de treinamento e de teste numa razão de 3:1 e mantivemos o equilíbrio entre as classes. Assim, treinamos cada classificador com os mesmos arquivos, 750 "positivos" e 750 "negativos". A seleção de atributos para representar os textos foi feita com uma abordagem do tipo Sacola de Palavras (do inglês, *Bag of Word — BoW*).

```
def set_feats(feats):
    negids = movie_reviews.fileids('neg')
    posids = movie_reviews.fileids('pos')

    negfeats = [(featx(movie_reviews.words(fileids=[f])), 'neg') for f in negids]
    posfeats = [(featx(movie_reviews.words(fileids=[f])), 'pos') for f in posids]

    negcutoff = len(negfeats)*3/4
    poscutoff = len(posfeats)*3/4

    train = negfeats[:negcutoff] + posfeats[:poscutoff]
    test = negfeats[negcutoff:] + posfeats[poscutoff:]

    return train, test
```

Figura 18. Código para criação dos conjuntos de treinamento e de teste

Procedido o treinamento, medimos a precisão e o tempo de processamento da classificação automática dos 500 arquivos do conjunto de teste. Os resultados obtidos foram os seguintes:

Quadro 1. Desempenho dos classificadores - EXPERIMENTO I

CLASSIFICADOR	PRECISÃO	TEMPO DE RESPOSTA (t_{resp})
Máxima Entropia	75%	$11 \text{ s} < t_{\text{resp}} < 16 \text{ s}$
Bayesiano	71%	$t_{\text{resp}} < 10 \text{ s}$
Árvore de Decisão	72%	$10 \text{ s} < t_{\text{resp}} < 14 \text{ s}$

Quanto à precisão, não observamos diferenças significativas nos resultados alcançados por cada classificador. Já em relação ao tempo necessário para realizar a classificação por completo, observamos que o classificador Bayesiano foi em média 20% mais rápido do que os demais.

4.4.1.2. EXPERIMENTO 2: INFLUÊNCIA DO ESPAÇO DE ATRIBUTOS

Para estudar a influência do tamanho e do tipo do espaço de atributos (*features*) sobre o desempenho dos classificadores, eliminamos e selecionamos alguns atributos e refizemos o experimento anterior.

Adotamos uma abordagem do tipo Sacola de Palavras, em que todas as palavras de uma mensagem são consideradas como unidades individuais de análise (*unigram*) (SEBASTIANI, 2002; GIORGINO, 2004), e selecionamos apenas os termos mais informativos (*High Information Words*) (MONTANES, 2003) a serem usados na representação dos arquivos contidos no *corpus* "Movie Reviews". Para isso, utilizamos recursos do NLTK e eliminamos artigos, preposições e pontuações (*StopWords*). O código da Figura 19 descreve esse procedimento.

```
from featx import high_information_words, bag_of_words_in_set
labels = EuTu.categories()
labeled_words = [(l, EuTu.words(categories=[l])) for l in labels]
high_info_words = set(high_information_words(labeled_words))
feat_det = lambda words: bag_of_words_in_set(words, high_info_words)
lfeats = label_feats_from_corpus(EuTu, feature_detector=feat_det)
train_feats, test_feats = split_label_feats(lfeats)
nb_classifier = NaiveBayesClassifier.train(train_feats)
accuracy(nb_classifier, test_feats)
```

Figura 19. Código para extrair as palavras mais significativas

Assim, foi possível manter o significado original dos textos e proporcionar dados com

mais qualidades para os processos de treinamento. Os resultados obtidos encontram-se resumidos no Quadro 2:

Quadro 2. Desempenho dos classificadores - EXPERIMENTO II

CLASSIFICADOR	PRECISÃO	TEMPO DE RESPOSTA (t_{resp})
Máxima Entropia	80%	$11\text{ s} < t_{resp} < 15\text{ s}$
Bayesiano	78%	$t_{resp} < 10\text{ s}$
Árvore de Decisão	78%	$9\text{ s} < t_{resp} < 12\text{ s}$

Mesmo com um aumento, as precisões dos três classificadores permaneceram próximas, sendo o classificador Bayesiano o que apresentou maior ganho.

Quanto ao tempo de resposta, não conseguimos identificar diferenças significativas, apesar de ter parecido que ocorreram compensações capazes de acelerar o processamento, com uma leve vantagem para o classificador Árvore de Decisão. Preferimos registrar esse fato, sem considerá-lo como incontestável. Achamos que ainda são necessários outros testes para chegar a uma conclusão definitiva sobre esse resultado.

Concluimos que a eliminação e seleção de atributos é realmente um passo importantíssimo no pré-processamento dos dados para a classificação. Com ela, é possível construir um espaço de atributos que aumenta a eficiência e a precisão dos classificadores.

4.4.1.3. A ESCOLHA DO CLASSIFICADOR

Alguns dos trabalhos que estudamos sobre classificação de documentos indicam que o algoritmo de Bayes é um dos métodos de aprendizagem de máquina mais populares e está presente nos melhores classificadores (KIAN, HAI, HWEE, 2002) (COLAS, BRAZDIL, 2006).

Nossos experimentos serviram para reforçar essa ideia. Eles mostraram que, além de simples de implementar e prático de usar, o classificador Bayesiano aparenta ter maior escalabilidade e menor tempo de processamento em relação aos classificadores Árvore de Decisão e de Máxima Entropia, o que, para o *Eu-Tu*, significa melhores condições para

proporcionar mediações oportunas.

Por tudo isso, verificamos haver evidências da razoabilidade do emprego do classificador Bayesiano para cumprir o objetivo de nossa pesquisa.

4.4.2. ETAPA II: CONSTRUÇÃO DO CLASSIFICADOR

Uma vez determinado o algoritmo de classificação a ser empregado, passamos à construção do classificador. Ainda utilizando o NLTK como ferramenta de Processamento de Linguagem Natural, realizamos um novo estudo piloto que envolveram outros dois aspectos relacionados aos parâmetros básicos do classificador.

Na prática, nem sempre será possível contar com um *corpus* de grande tamanho para treinamento em todos os domínios considerados. A literatura mostra que para o classificador Bayesiano binário, o conjunto de treinamento não pode ser muito pequeno. Caso contrário, o desempenho tenderá a ser baixo.

Outro aspecto considerado foi que, até então, havíamos testado apenas classificadores binários. Para a aplicação do Modelo de Avaliação proposto (LOPES, 2007), era necessário considerar um número maior de categorias e descobrir se um único classificador Bayesiano multiclasse seria uma boa estratégia de classificação a ser aplicada. Vimos que a combinação (*ensemble*) de classificadores Bayesianos binários individuais (DIETTERICH, 2000) apresenta melhores resultados do que um único classificador multiclasse. Nessas circunstâncias, chegamos à solução final proposta para esta pesquisa.

4.4.2.1. EXPERIMENTO III: INFLUÊNCIA DO TAMANHO DO CONJUNTO DE TREINAMENTO

Este experimento teve como objetivo verificar especificamente como a precisão do classificador Bayesiano é afetada pelo tamanho desse conjunto. Idealmente, deveríamos também ter considerado esse aspecto para a escolha do classificador a ser empregado e incluído os classificadores de Máxima Entropia e Árvore de Decisão para uma análise mais precisa. Entretanto, como os estudos iniciais e os resultados obtidos nos experimentos anteriores já indicavam que o classificador Bayesiano apresentava características favoráveis para ser aplicado a esta pesquisa, diante da premência de tempo para prosseguirmos com os trabalhos, consideramos estudar os efeitos provocados pelo tamanho do conjunto de

treinamento apenas para esse tipo de classificador. Sabíamos que tal decisão poderia afetar os resultados que fossem encontrados, mas conjecturamos que, se isso ocorresse, seria minimamente, o que poderia ser verificado em trabalhos futuros.

Refizemos o Experimento I para o classificador Bayesiano, porém alterando a quantidades de arquivos do conjunto de treinamento. Ao todo, foram realizadas 6 interações. Em cada uma delas, novos arquivos para treinamento eram adicionados, mantendo-se o equilíbrio entre as duas classes, conforme indicado no Quadro 3.

Quadro 3. Variação do tamanho do conjunto de treinamento - EXPERIMENTO III

Interação	Conjunto de Treinamento	Quantidade de Arquivos
1	movie_100p	1000 (500 pos e 500 neg)
2	movie_90p	900 (450 pos e 450 neg)
3	movie_75p	750 (375 pos e 375 neg)
4	movie_50p	500 (250 pos e 250 neg)
5	movie_25p	250 (125 pos e 125 neg)
6	movie_10p	100 (50 pos e 50 neg)

Os resultados obtidos encontram-se representados no gráfico da Figura 20. Verifica-se que o conjunto de treinamento não pode ser muito pequeno, uma vez que a precisão do classificador Bayesiano é influenciada pela quantidade de documentos desse conjunto. Apesar de ser uma observação já prevista por outros trabalhos de pesquisa (SHI Yong-feng; ZHAO Yan-ping, 2004; BAHARUDIN, LEE, KHAN, 2010) e, de certa forma, até intuitiva, nossos resultados identificaram a tendência de uma de relação linear entre a precisão e o tamanho do conjunto de treinamento, dentro do intervalo 100-1000 arquivos do nosso estudo. Ainda assim, é interessante que sejam feitos mais estudos semelhantes a esse para verificar se essa relação persiste em outros intervalos.

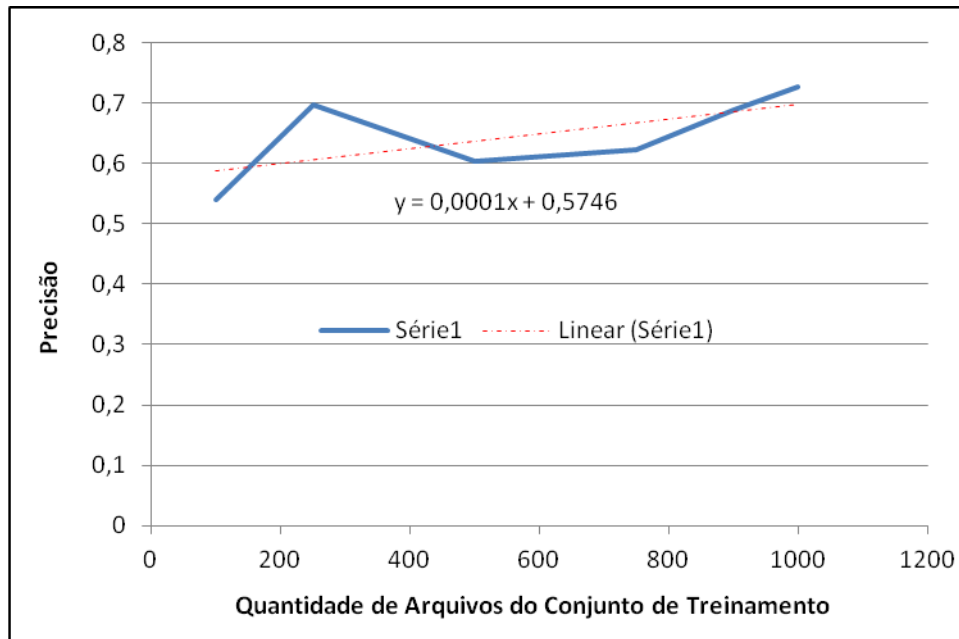


Figura 20. Influência do tamanho do conjunto de treinamento no desempenho do classificador Bayesiano binário - EXPERIMENTO III

Como, na prática, essa é uma condição nem sempre possível de satisfazer, o modelo iterativo do *Eu-Tu* tem um papel fundamental, permitindo que sejam realizados treinamentos continuados, o que faz o classificador aprender novos dados a cada novo ciclo. Dessa forma, ainda que a classificação tenha baixa precisão de início, será aperfeiçoada gradativamente.

4.4.2.2. EXPERIMENTO IV: INFLUÊNCIA DO NÚMERO DE CATEGORIAS

Este experimento teve como objetivo verificar como a precisão do classificador Bayesiano é afetada pelo número de categorias consideradas.

Ressalta-se que, de maneira intuitiva, é bastante simples admitir que existe uma relação inversa entre essas duas grandezas, ou seja, quanto maior o número de categorias, menor será a precisão. Logo, pode-se concluir com facilidade que, dentre os classificadores multiclases, o classificador binário é o de maior precisão. Existem trabalhos na literatura científica que atestam isso, o que, de certa forma, tornaria desnecessário a realização de mais um teste. Entretanto, tínhamos também como interesse verificar se haveria a possibilidade de aplicar um único classificador Bayesiano multiclasse de 11 categorias para o Modelo de Avaliação que havíamos escolhido.

Construímos *corpora* de treinamento e teste, usando a mesma base de dados empregada por Lopes (2007) em sua pesquisa. Essa base de dados foi formada por mensagens

de um fórum categorizado, implementado na plataforma *Pii* (ELIA, SAMPAIO, 2001) para um curso de pós-graduação em Coordenação Pedagógica, do Exército Brasileiro, realizado em sua maior parte na modalidade de EAD/TIC. Sobre os testes feitos com essa base de dados, Lopes (2007) descreve:

"A ferramenta foi apropriada como fórum de discussão, onde um tema é proposto para todos os alunos. Cada um escreve seu comentário, tornando-se disponíveis para todos os participantes; portanto, o participante pode ler não só sua opinião, mas também a de outros alunos, como se estivesse discutindo o assunto em sala de aula. O professor acompanha e orienta o fórum, propondo novos rumos para a discussão e esclarecendo pontos. A ferramenta ofereceu apoio tecnológico para que os debatedores pudessem discutir a partir de seus computadores pessoais ou do trabalho e em momentos de disponibilidade individual. Permitiu a organização em grupos e o devido acompanhamento pelo professor, já que o ambiente disponibiliza os *Logs* das interações de forma cronológica ou por debatedor, além de estatísticas para posterior análise."

Nota-se que o objetivo de Lopes (2007) foi extrair informações das mensagens trocadas por grupos de alunos em uma atividade colaborativa, para avaliar a aprendizagem proporcionada pelas discussões que ocorreram. Além de representarem conteúdos semelhantes aos que poderão ser manipulados, na prática, pelo *Eu-Tu*, esse conjunto de mensagens também já havia sido pré-classificado com os mesmos indicadores selecionados para nossa pesquisa, constituindo assim uma boa referência para testar o desempenho do protótipo do classificador que construímos.

Quadro 4. Distribuição das mensagens por categorias - EXPERIMENTO IV

CATEGORIA	QUANTIDADE DE MENSAGENS
Saudação	13
Debate	9
Motivação	5
Social	5
Informação	4
Confirmação	3
Negação	3
Tarefa	3
Esclarecimento	2
Indagação	2
Agradecimento	1

Originalmente, existiam 50 mensagens distribuídas em um corpus com 11 categorias distintas. Diferente dos experimentos anteriores, essa distribuição não era uniforme, como mostra o Quadro 4.

Com base nos resultados do EXPERIMENTO III, percebemos que, se realizássemos as classificações diretamente a partir desse *corpus*, a precisão seria minimizada, por causa do pequeno número de mensagens, tanto no total quanto por categorias.

Para minimizar o desequilíbrio entre as categorias adotamos um procedimento iterativo de criação de novos corpora por meio da partição do corpus original. De início, ordenamos as categorias de acordo com suas quantidades de mensagens.

A partir daí, criamos um corpus com apenas duas classes, uma constituída pela categoria com maior número de mensagens e outra denominada "Negativa", que reuniu todas as outras 10 categorias restantes.

No Quadro 4, podemos ver que, para esse primeiro *corpus*, as duas classes foram "Saudação", com 13 mensagens e "Negativa", com 37 mensagens.

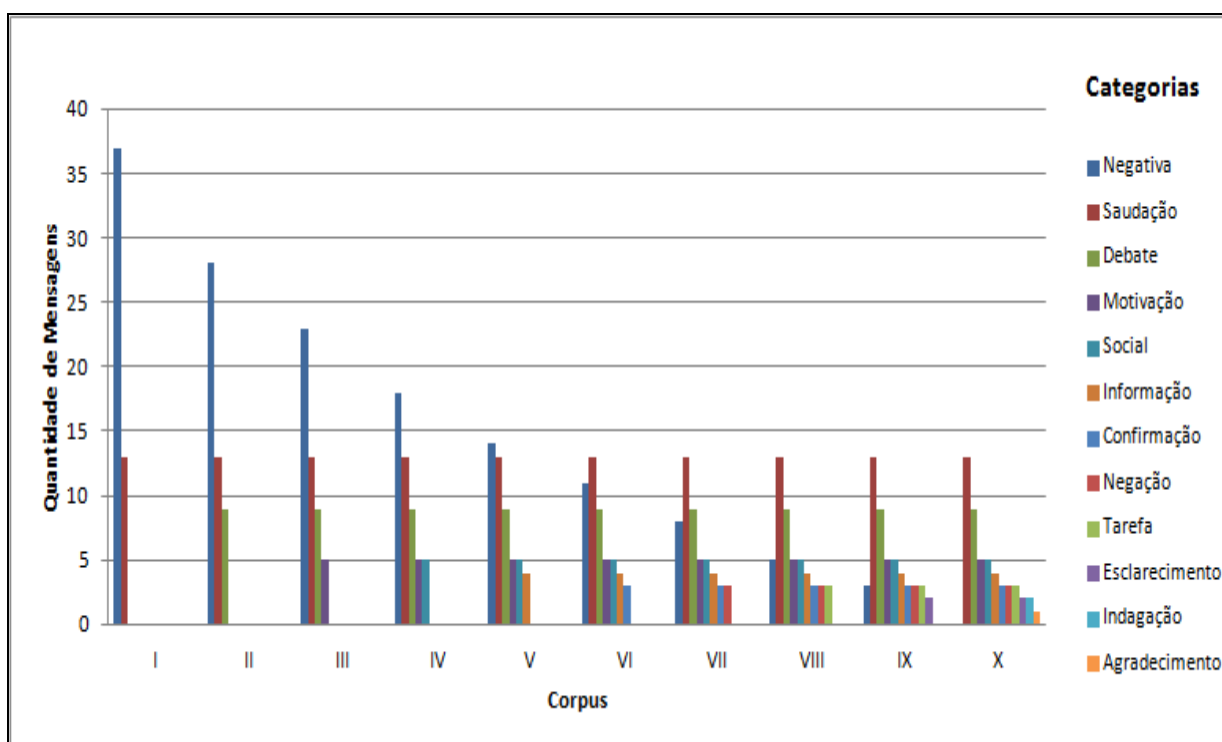


Figura 21. Distribuição das categorias - EXPERIMENTO IV

A cada nova iteração, separamos a maior categoria componente da classe "Negativa" para formar uma classe independente e dar origem a um novo *corpus*. Esse procedimento foi

repetido até alcançarmos o *corpus* original, com 11 categorias. A Figura 21 mostra como ficou a distribuição das categorias nos *corpora* criados.

Para cada *corpus* criado, formamos conjuntos de treinamento e de teste na proporção de 3:1, semelhante ao EXPERIMENTO I, e desenvolvemos um novo classificador Bayesiano multiclasse, usando ferramentas do NLTK, para classificar as mensagens. Com os resultados obtidos, construímos o gráfico da Figura 22.

Esse gráfico comprova que a precisão de um classificador Bayesiano multiclasse decai com o aumento do número de categorias, sendo os binários os mais precisos.

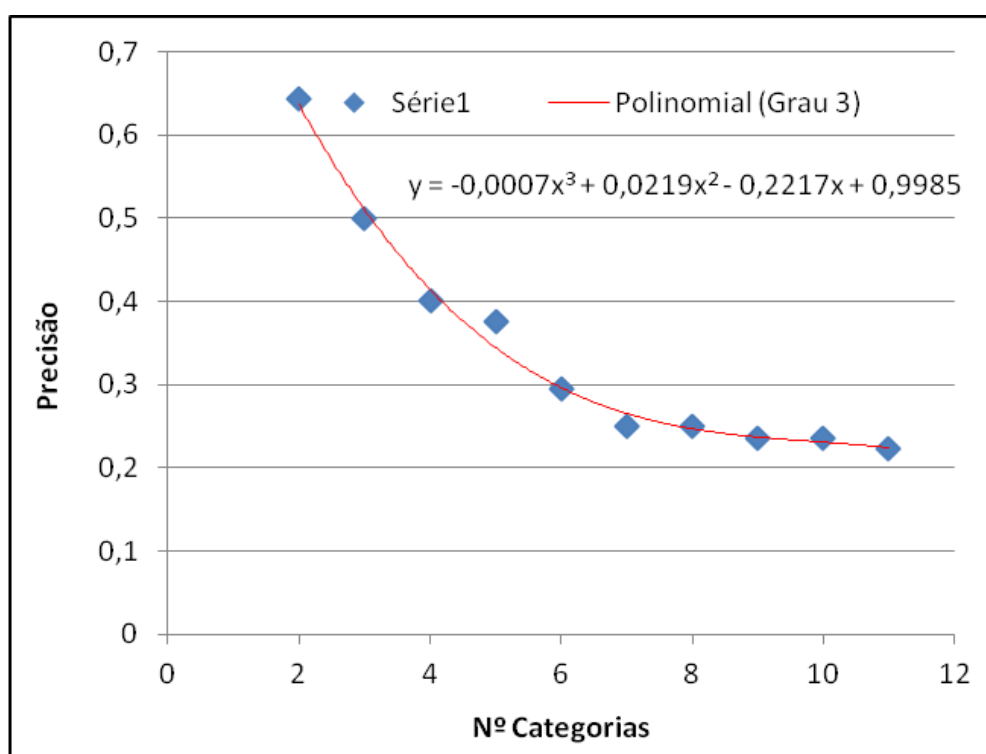


Figura 22. Influência do número de categorias na precisão do classificador Bayesiano multiclasse – EXPERIMENTO IV

Como almejávamos, identificamos que, para um classificador de 11 categorias, a precisão fica em torno de 22%, um valor muito baixo para ser usado em avaliações. Portanto, o emprego desse tipo de classificador multiclasse inviabiliza o Modelo de Avaliação que adotamos.

Nossa alternativa foi aplicar um conjunto (*ensemble*) de classificadores binários para solucionar esse problema. No próximo capítulo, descrevemos como essa solução foi validada para a construção de um protótipo do Módulo de Avaliação do *Eu-Tu*.

CAPÍTULO 5: RESULTADOS E ANÁLISE

Os resultados dos experimentos que realizamos revelaram que um único classificador Bayesiano multiclasse não proporciona precisão suficiente para ser empregado em nossa pesquisa. Com isso, decidimos aplicar um conjunto (*ensemble*) de classificadores binários como estratégia de classificação.

A fim de validar essa solução, realizamos um teste com um protótipo que desenvolvemos. Procuramos comprovar que a precisão oriunda da nova estratégia de classificação permite o emprego do Modelo de Avaliação adotado, com todas as suas 11 categorias.

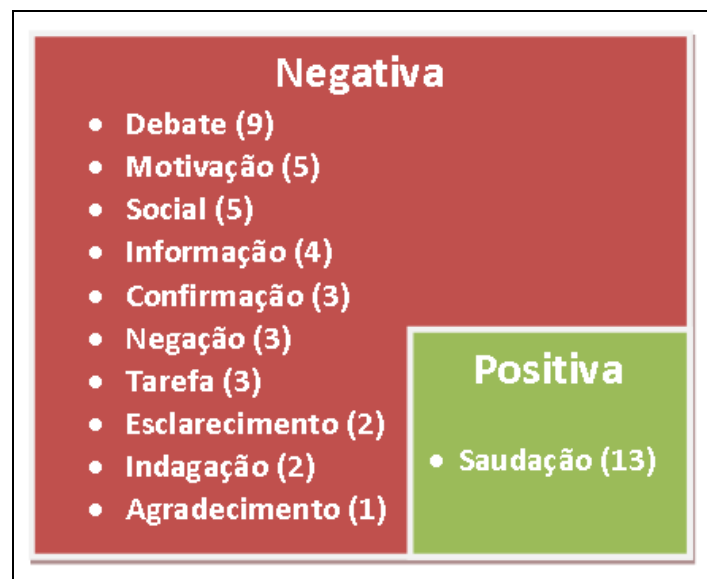


Figura 23. Particionamento usado para criar o *corpus* do classificador da categoria "Saudação"

Em primeiro lugar, foi necessário criar e treinar um classificador específico para cada categoria. Com essa finalidade, formamos novos *corpora* particionando apropriadamente o *corpus* original do EXPERIMENTO IV (Seção 4.4.2.2), de modo que uma categoria pudesse ser considerada como "alvo" (Positiva) e as demais, como "não-alvo" (Negativa). A Figura 23 ilustra o particionamento tendo a categoria "Saudação" como "alvo". Os valores entre parênteses representam as quantidades de mensagens existentes em cada categoria.

Uma vez tendo criado e treinado separadamente os 11 classificadores binários, foi possível reuni-los em um único conjunto usando novamente o NLTK, conforme mostra a Figura 24.

```
>>> trainf = lambda train_feats: NaiveBayesClassifier.train(train_feats)
labelset = set(EuTu.categories())
classifiers = train_binary_classifiers(trainf, multi_train_feats, labelset)
len(classifiers)

multi_classifier = MultiBinaryClassifier(*classifiers.items())
```

Figura 24. Código para construção do conjunto de classificadores binários

Para implementar a estratégia de classificação, empregamos uma abordagem de votação tipo "um contra o resto", em que todos os classificadores apresentam seus votos (classificação) para cada mensagem, conforme descreve o código da Figura 25.

```
multi_classifier.classify(featsG)

for k in classifiers.keys():
    print classifiers[k].classify(featsG)
```

Figura 25. Código para determinar o voto dos classificadores

Se apenas um voto for "positivo", a mensagem é considerada da mesma categoria de seu classificador. Caso contrário, se houver mais de um voto, é considerada da categoria do classificador de maior precisão que votou positivamente. Se não houver voto algum, é classificada como "desconhecida". O diagrama da Figura 26 ilustra esse mecanismo de votação.

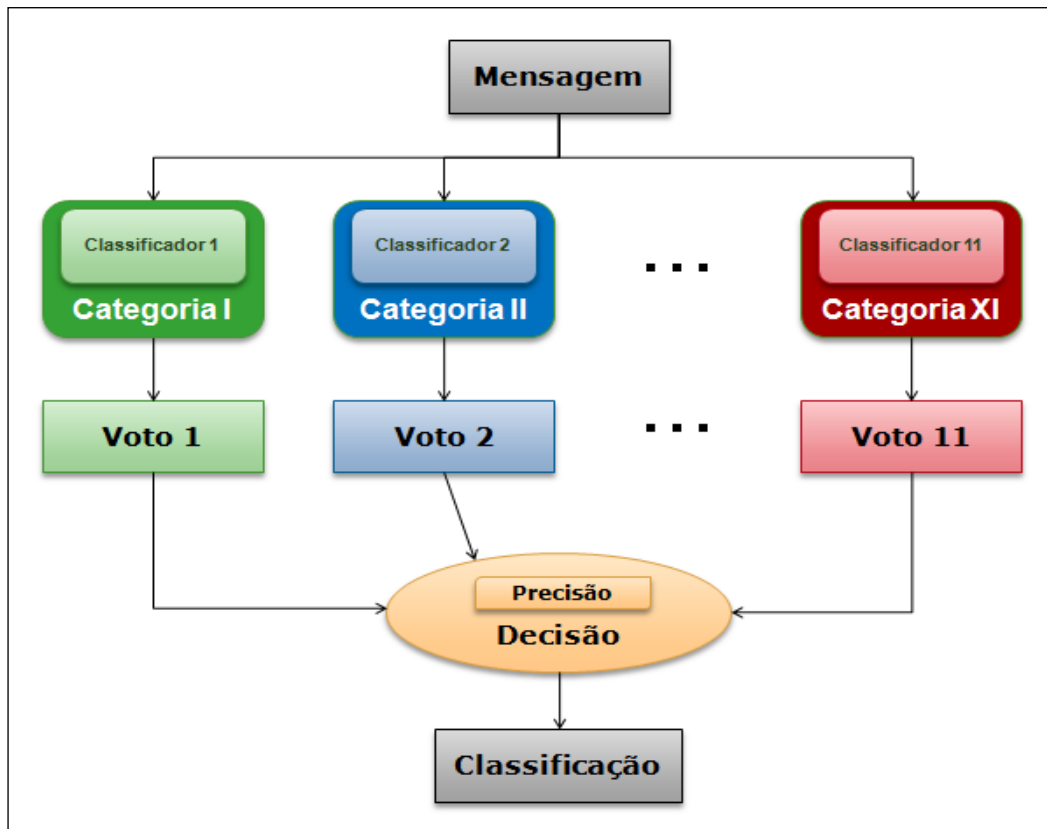


Figura 26. Diagrama do esquema de votação para o conjunto de classificadores binários

O código da Figura 27 descreve como calculamos a precisão de cada classificador.

```

for k in multi_classifier.labels():
    print '[%s]: ' % k + ' Precisão: ' + repr(nb_precisions[k])
    + ' Cobertura: ' + repr(nb_recalls[k])

for k in classifiers.keys():
    probs = classifiers[k].prob_classify(featsG)
    print 'probs.samples para %s : ' % k
    probs.samples()
    print classifiers[k].classify(featsG)
    p, n = probs.samples()
    print 'probs.prob(%s): ' % p
    probs.prob(p)
    print 'probs.prob(%s): ' % n
    probs.prob(n)
    print
  
```

Figura 27. Código para determinar a precisão individual dos classificadores binários

Como resultado da implementação da nova estratégia de classificação, a precisão da combinação de 11 classificadores ficou em torno de 40%. Esse valor equivale a de um

classificador Bayesiano multiclasse de apenas 5 categorias, comprovando que, para nossos objetivos, a combinação de classificadores Bayesianos binários é realmente uma estratégia mais adequada do que um único classificador multiclasse.

Além do mais, a precisão alcançada permite que o conjunto de classificadores desenvolvido seja usado como uma ferramenta auxiliar para a avaliação, reduzindo os esforços despendidos pelos professores para a correção de inúmeras tarefas, principalmente em cenários de EAD, que costumam apresentar grandes quantidades de alunos. Tal fato permite cumprir o objetivo geral inicialmente proposto para este trabalho, firmando convicção a respeito da veracidade da hipótese estabelecida.

Quanto ao objetivo específico, pôde-se perceber que, tecnicamente, é possível alcançar a precisão necessária (60% — LUI, LI, CHOY, 2007) para que o conjunto de classificadores seja usado como uma ferramenta automática de avaliação do progresso da aprendizagem em fóruns eletrônicos de discussões.

Semelhante ao que ocorreu no EXPERIMENTO IV, notamos uma redução na precisão individual dos classificadores provocada pelo desbalanceamento de classes. Na abordagem "um contra o resto", as categorias com menos mensagens tendem a cometer mais erros, conforme visto no EXPERIMENTO III.

Lamentavelmente, não conseguimos tratar desse problema ainda neste trabalho, devido à redução de tempo provocada por escolhas e reorientações de caminhos imprevistas, que tiveram que ser feitas no decorrer da pesquisa. Acreditamos que, se esse desbalanceamento for tratado adequadamente, a precisão será ainda maior (KOTSIANTIS, KANELLOPOULOS, PINTELAS, 2006; GARCIA et al, 2007).

Verificamos que uma possibilidade é modificar a amostragem das classes para que haja um equilíbrio entre elas, seja aumentando o número de elementos das classes minoritárias (*oversampling*) ou reduzindo o das majoritárias (*undersampling*).

Hwanjo Yu, Chengxiang Zhai e Jiawei Han (2003) apresentaram um método interessante de *oversampling* que utiliza textos ainda não classificados para a construção de um classificador. Considerando a grande quantidade de informações textuais disponíveis, esse método reduz o esforço necessário para pré-classificar mensagens de treinamento e amplia o número de fontes de dados.

Um método de *undersampling* bastante usado é o *boosting*. Zhipeng Xie (2009) reporta que encontrou bons resultados ao aplicar esse método a classificadores Bayesianos.

Dentre os algoritmos de *boosting* mais conhecidos, encontra-se o *Adaboosting* (FERREIRA, 2007). Duarte (2009) generalizou esse algoritmo, permitindo a utilização de uma distribuição arbitrária para a distribuição dos pesos dos exemplares classificados. Xu-Ying Liu, Jianxin Wu e Zhi-Hua Zhou (2006) apresentaram outra versão modificada de *boosting* também capaz de reduzir o desperdício de informações provocado pelo descarte de exemplares das classes majoritárias.

CAPÍTULO 6: CONCLUSÕES E TRABALHOS FUTUROS

6.1. VISÃO GERAL

Em um mundo complexo, as possibilidades de ensinar e aprender foram ampliadas. O avanço do desenvolvimento tecnológico, ocorrido principalmente nos últimos 50 anos, nos deu facilidades para compartilhar ideias e pensamentos, algo que, outrora, era restrito a alguns.

Professores e alunos se tornaram igualmente fontes de conhecimentos que não são transferidos de uma mente para outra, como se fossem algo concreto. Ideias passaram a ter que ser pensadas e experiências, revividas, trazendo alternativas para que o ensino de um mesmo assunto à pessoas diferentes deixe de ser um obstáculo quase que intransponível para as práticas educacionais.

Com tantas mudanças ocorrendo o tempo todo, ensinar e aprender precisam ser atividades simples, precisas e eficientes, para que também sejam praticadas constantemente.

Nesse contexto, as interações dos indivíduos destacam-se como o ponto de partida para a construção da nova realidade social, dando origem a uma nova forma de ensinar e aprender que chamamos de “processo de ensino e aprendizado centrado nas interações”.

A presente pesquisa teve como objetivo geral, explorar uma das possíveis maneiras de promover esse processo, através de uma forma indireta de avaliar e mediar interações, que se apoia na discussão escrita em um fórum eletrônico. Acreditamos que discussões efetivas contribuem para o desenvolvimento de novas percepções, conhecimentos e perspectivas.

Propomos a criação de um sistema informatizado, que denominamos de *Eu-Tu*, com a expectativa de que seu emprego servirá para dar suporte às diferenças entre seus usuários e tornar as ações pedagógicas mais adequadas para tratar os efeitos da singularidade tecnológica.

O *Eu-Tu* possui um modelo simples e iterativo, composto basicamente por dois módulos: um de avaliação e outro de mediação. O Módulo de Avaliação tem por finalidade verificar a qualidade das interações e gerar indicadores para o Módulo de Mediação, através

da análise do conteúdo de mensagens. O Módulo de Mediação é responsável por delinear e implementar ações que estimulem interações efetivas (Estratégias de Mediação), a partir dos indicadores gerados pelo Módulo de Avaliação.

Especificamente, direcionamos este trabalho para a construção do Módulo de Avaliação, como forma de dar início ao desenvolvimento do *Eu-Tu*.

6.2. PROBLEMAS ENCONTRADOS

Alguns estudos - necessários e importantes, mas não previstos – aliados às escolhas e reorientações de caminhos que tiveram que ser feitas ao longo do estudo demandaram bastante tempo, acarretando atrasos na construção e validação do modelo proposto para o *Eu-Tu*.

Durante a construção do Módulo de Avaliação, encontramos alguns desafios técnicos que serviram para dar direção à solução proposta nesta pesquisa. Para tratar desses desafios, empregamos técnicas de Processamento de Linguagem Natural e realizamos experimentos utilizando um conjunto de ferramentas *open source*, escritas em *Python*, chamado de *Natural Language Toolkit* – NLTK.

Analizamos empiricamente o desempenho de três classificadores binários: Bayesiano, Árvore de Decisão e Máxima Entropia. Consideramos a precisão e o tempo necessário para realizar a classificação por completo, como medidas para avaliação. Quanto à precisão, não observamos diferenças significativas. Já em relação ao o tempo de processamento, observamos que o classificador Bayesiano foi relativamente o mais rápido.

Estudamos a influência do tamanho e do tipo do espaço de atributos (*features*) sobre o desempenho dos classificadores. Percebemos que com a eliminação e seleção de atributos é possível construir um espaço de atributos que aumenta a eficiência e a precisão dos classificadores. Descobrimos que dos três classificadores estudados, o Bayesiano é o que mais se beneficia com a redução do número de atributos.

Analizamos como a precisão do classificador Bayesiano binário é afetada pelo tamanho do conjunto de treinamento. Notamos que, como tendência, existe uma relação linear entre essas duas grandezas. Isso sugere que para se ter uma boa precisão, o conjunto de treinamento não pode ser muito pequeno.

Também estudamos como a variação do número de categorias altera a precisão do

classificador Bayesiano multiclasse. Vimos que essa precisão decai com o aumento do número de categorias.

Descobrimos que a combinação de classificadores Bayesianos binários consegue apresentar melhores resultados do que um único classificador multiclasse. Por exemplo, no nosso estudo vimos que utilizando um *ensemble* com número N ($=11$) de classificadores binários e um *corpus* de teste com N ($=11$) categorias, foi possível alcançar uma precisão equivalente (no caso 40%) a que seria alcançada usando um classificador Bayesiano multiclasse com apenas aproximadamente $N/2$ (≈ 5) categorias, sendo tudo mais igual no processo de comparação. Naturalmente, seria importante fazer outros estudos no sentido de verificar como este ganho se comporta em função do número inicial de categorias do *corpus* de teste.

Vimos ainda que para a combinação de classificadores, existe um efeito provocado pelo desbalanceamento de classes que reduz a precisão global quando se usa uma estratégia de votação do tipo "um contra o resto", onde cada classificador individualmente é responsável por distinguir uma única categoria dentre as demais. Se esse desbalanceamento for tratado adequadamente, a precisão do *ensemble* pode ser ampliada.

Apesar de algumas limitações e simplificações, os resultados alcançados deram suporte à consecução do objetivo da pesquisa e à hipótese formulada.

6.3. CONTRIBUIÇÕES

Conseguimos construir um conjunto (*ensemble*) de classificadores Bayesiano binários, capaz de classificar automaticamente textos, com base no Modelo de Análise (taxonomia) desenvolvido por Lopes (2007), que considera a construção do conhecimento como indicador de aprendizagem.

Por meio dos estudos e testes feitos, aprendemos como utilizar diversos algoritmos, plataformas e software necessários para a construção do *Eu-Tu*, incluindo a implementação de rotinas e programas que serão reutilizados e de processos para otimizar o classificador automático.

Conseguimos criar um instrumento que pode auxiliar a avaliação continuada do ensino e da aprendizagem. Além de contribuir para avaliações mais justas, o mecanismo automático desenvolvido contribui para reduzir os esforços despendidos na correção de inúmeras tarefas.

Essa é uma grande vantagem, principalmente para cursos em ambientes de EAD, onde se costuma ter grandes quantidades de alunos sobre a batuta de um único tutor.

Duas observações importantes, que também podem auxiliar outras pesquisas, surgiram no decorrer da realização dos experimentos deste trabalho.

A primeira é que a ideia de um classificador universal, que engloba todos os domínios do conhecimento é quase que utópica. Classificar é uma tarefa envolvida pela subjetividade. Classificar mensagens parece ser ainda mais, devido à existência de variáveis latentes escondidas nos conteúdos dos textos. Para encontrá-las, é preciso considerar o contexto no qual as mensagens foram escritas, uma vez que os esquemas de codificação são tipicamente construídos dependentes desse contexto.

Portanto, é mais plausível pensar em classificadores especializados que podem ser aperfeiçoados, à medida que forem aprendendo novos domínios. Nesse aspecto, o *Eu-Tu*, por ser um sistema automático e iterativo, em que os resultados gerados por múltiplas perspectivas de análise podem servir como dados para novos processamentos, permite a criação de classificadores otimizados para diversos domínios.

A segunda é que a autoria da classificação das mensagens tem influência direta sobre a confiabilidade de um *corpus*. Surpreendentemente, vimos que o conteúdo de mensagens pré-classificadas por seus autores, nem sempre representam o significado da pós-classificação que receberam, seja pelo classificador automático, ou seja, por terceiros¹⁷. Notamos que há uma diferença entre o que intencionavam escrever para o que realmente escreveram. Essa falta de correspondência resulta na formação de *corpora* distanciados da verdade. Para o modelo ora proposto neste trabalho, isso significaria enxergar construção de conhecimento, quando, de fato, não houve, e vice-versa.

Notamos também que a classificação feita por um especialista no assunto tratado é a que mais se alinha com o conteúdo das mensagens. Isso indica que, para um classificador automático gerar índices que permitam analisar o processo de ensino e aprendizagem fidedignamente, o *corpus* mais adequado a ser usado para treinamento e teste é aquele formado por mensagens pré-classificadas por um professor ou mediador.

¹⁷ Foi pedido ao professor da disciplina que também fizesse uma pós-classificação das mesmas mensagens.

6.4. TRABALHOS FUTUROS

A análise de interações por mensagens escritas engloba diversas áreas de estudo, com inúmeras alternativas a serem exploradas. Neste trabalho, demos apenas o passo inicial para o desenvolvimento do *Eu-Tu*. Sabemos que isso representa apenas "uma semente" lançada em um "vasto e fértil terreno" para o ensino e aprendizagem centrados em interações. E, para que essa "semente" dê muitos frutos, novos trabalhos são necessários.

Achamos que, antes de construir o Módulo de Mediação, seja necessário incluir uma etapa de testes com comunidades reais de alunos para aperfeiçoar a eficiência da do Módulo de Avaliação.

Pretendemos explorar novas técnicas para classificação automática, como, por exemplo, a otimização por colônia de formigas (JIAO, FENG, 2010), o uso da estrutura dialógica (SU NAM KIM, CAVEDON, BALDWIN, 2010) e ontologias (KAO, POTEET, 2007). Especificamente para o problema do desbalanceamento, Kotsiantis e Pintelas (2003) propuseram o uso de agentes especialistas.

Outra alternativa é empregar técnicas que selecionem os atributos mais significativos para a representação dos textos. Forman e Kirshenbaum (2008) apresentaram uma técnica eficiente de extração de atributos. Baharudin, Lee e Khan (2010) propuseram o uso de algoritmos genéticos para extração desses atributos. Outros trabalhos sugerem o uso da informação lexical, semântica, estrutural e seqüencial (WIEBE et al, 2004; JINGNIAN et al, 2009; SU KIM et al, 2011; FITZGERALD et al, 2011). Yeh, Hirschman e Morgan (2003) afirmam que as técnicas estatísticas não são suficientes para a mineração de dados e que classificações mais precisas são realizadas quando se consideram os aspectos semânticos das mensagens. Alguns trabalhos sugerem que a classificação baseada em ontologias é capaz de captar mais precisamente aspectos relacionados ao domínio das mensagens (YEH, HIRSCHMAN, MORGAN, 2003; SONG, et al, 2005; TENENBAUM, SHAPIRA, SHOVAL, 2008). Yi-Hsing e Hsiu-Yi (2008) inclusive desenvolveram um sistema automático de classificação de documentos com base em ontologias e no classificador de Bayes.

Outro importante passo a ser dado está relacionado com a aplicação prática do *Eu-Tu*. Para analisar outras dimensões do processo de ensino e aprendizagem, que considerem aspectos como conativos, emocionais e sociais (DAY, CARROLL, 2004), é necessário descobrir qual a precisão mínima que o classificador deve ter para se aplicar um determinado

Modelo de Avaliação. Uma forma de se descobrir essa precisão é através de estudos comparativos entre a classificação feita manualmente por especialistas humanos e pelo classificador automático com base em diferentes Modelos de Avaliação, como o realizado por Hale, Jorgenson e Gamble (2011).

Assumimos que a combinação social associada à experiência do professor é uma forma de mediação efetiva, capaz de desenvolver as habilidades necessárias à inovação. É necessário verificar se isso realmente ocorrerá com a interação de parceiros de aprendizagem adequados. Nesse aspecto, a realimentação do sistema tem papel fundamental. A conjunção de novos e velhos resultados servirá para criar regras que poderão levar à combinação otimizada. Assim, à medida que o sistema for sendo usado, também poderá ser aperfeiçoado.

Também é preciso descobrir como dar suporte ao professor para orquestrar as discussões através de suas mensagens, de modo que haja orientações adequadas quando ocorrerem problemas e encorajamento, quando as conversações forem produtivas.

6.5. CONSIDERAÇÕES FINAIS

Neste trabalho, referenciais filosóficos, pedagógicos e tecnológicos foram considerados para desenvolver uma ferramenta computacional capaz de aperfeiçoar o processo de ensino e aprendizagem centrado nas interações.

Engendramos um modelo onde esse aperfeiçoamento ocorre por dois processos, um que realiza ações e outro que as avaliam, conjugados de maneira que haja realimentação adequada para fazê-los convergir em resultados positivos. Não se trata de um modelo novo, mas lida com inovação. É simples, e trata da complexidade. Trabalha o coletivo, para potencializar o individual.

Instanciamos esse modelo e construímos um instrumento de classificação automática de textos, que mesmo com algumas restrições, sua aplicação ultrapassa os limites dessa pesquisa, permitindo que outros caminhos sejam explorados; alguns, inclusive, ainda não conhecidos.

REFERÊNCIAS

- ADAMS, P. **Grouped: How small groups of friends are the key to influence on the social web**. Berkeley, CA, USA: New Riders Press, 2011. 168 p. ISBN-10: 0321804112.
- ÁDAN COELLO, J. M.; FARIA, E. S. J.; MENEZES, W. S.; TOBAR, C. M. Conflito Sócio-cognitivo e Estilos de Aprendizagem na Formação de Grupos para o Aprendizado Colaborativo de Programação de Computadores. **Revista Brasileira de Informática na Educação**, v. 16, n. 3, p. 9-19, 2008.
- AGHDAM, M. H.; GHASEM-AGHAEI, N.; BASIRI, M. E. Text feature selection using ant colony optimization. **Journal Expert Systems with Applications**, Tarrytown, NY, USA, v. 36, n. 3, p. 6843–6853, April, 2009.
- AL HAMAD, A.; YACOB, N.; AL-ZOUBI, Y. Integrating ‘LearningStyle’ Information into Personalized e-Learning System. **IEEE Multidisciplinary Education Magazine**, v. 3, n. 1, p. 2-6, 2008.
- ALFONSECA, E.; CARRO, R. M.; MARTÍN, E.; ORTIGOSA, A.; PAREDES, P. The impact of learning styles on student grouping for collaborative learning: a case study. **Journal User Modeling and User-Adapted Interaction**, Hingham, MA, USA, v. 16, n. 3-4, p. 377 - 401, September, 2006.
- ALMEIDA, P. **Questões dos alunos e estilos de aprendizagem: um estudo com um público de ciências no ensino universitário**. Tese (Doutorado) — Departamento de Didática e Tecnologia Educativa, Universidade de Aveiro, Portugal, 2007.
- ALMEIDA, R. Q. **Perguntas e Respostas: Rau-Tu Sistema Colaborativo de Perguntas e Respostas**. Universidade Estadual de Campinas, Campinas, SP, Abril, 2006. Disponível em: <http://www.rau-tu.unicamp.br/doc/o-sistema-rau-tu-de-perguntas-e-respostas.php>. Acesso em: 20 Fev. 2012.
- ASTERHAN, C. S. C. Online moderation of synchronous e-argumentation. **International Journal of Computer-Supported Collaborative Learning**, v. 5, n. 3, p. 259–282, September, 2010.
- ASTERHAN, C. S. C.; EISENMANN, T. Online and face-to-face discussions in the classroom: A study on the experiences of ‘active’ and ‘silent’ students. In: O’MALLEY, C.; SUTHERS, D.; REIMANN, P.; DIMITRACOPOULOU, A. (Eds.). **International Conference on Computer Supported Collaborative Learning – CSCL’09**, 9., 2009, Rhodes, Greece. **Proceedings ...** Rhodes, Greece: International Society of the Learning Sciences, 2009. p. 132–136. ISBN: 978-1-4092-8598-4.
- BAHARUDIN, B.; LEE, L.H.; KHAN, K. A Review of Machine Learning Algorithms for Text-Documents Classification. **Journal of Advances in Information Technology**, v. 1, n. 1, p. 4-20, February, 2010. DOI:10.4304/jait.1.1.4-20.

- BAHRAMI, B. et al. Optimally interacting minds. **Science**, New York, N.Y., v. 329, n. 5995, p. 1081-1085. August, 2010. DOI: 10.1126/science.1185718.
- BANDURA, A. Social cognitive theory. In: VASTA, R. (Ed.). **Annals of Child: Six theories of Child Development**. Greenwich, CT: JAI Press, 1989., v. 6, p. 1-60. ISBN: 0892329688.
- BARCELLOS, C. D. et al. Sistema de Recomendação Acadêmico para Apoio a Aprendizagem. **RENOTE - Revista Novas Tecnologias na Educação**. CINTED-UFRGS, [Porto Alegre], v. 5, n. 2, Dezembro, 2007. ISSN 1679-1916.
- BARIANI, I. C. **Estilos Cognitivos de Universitários e Iniciação Científica**. Tese (Doutorado) – Faculdade de Educação, Universidade Estadual de Campinas, Campinas, 1998. Disponível em: <http://libdigi.unicamp.br/document/?code=vtls000125169>. Acesso em: 20 Fev. 2012.
- BARILLI, E. C. V. C. Avaliação: acima de tudo uma questão de opção. In: SILVA, M.; SANTOS, E. (Org.). **Avaliação da Aprendizagem em Educação Online**. São Paulo: Loyola, 2006. p. 153-170.
- BARNES, K., MARATEO, R.C.; FERRIS, P. Teaching and learning with the Net Generation. **Innovate Journal of Online Education**, v. 3, n. 4. April/May, 2007. Disponível em: http://innovateonline.info/pdf/vol3_issue4/Teaching_and_Learning_with_the_Net_Generation.pdf. Acessado em 20 Fev 2012.
- BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python**. Sebastopol: O'Reilly Media, Inc, 2009. 504 p. ISBN: 978-0-596-51649-9.
- BONABEAU, E. Editor's Introduction: Stigmergy. **Artificial Life**, Massachusetts, v. 5, n. 2, p. 95-96, Spring, 1999.
- BROWN, E. **The use of learning styles in adaptive hypermedia**. Tese (Doutorado) – Faculty of Science, School of Computer Science, University of Nottingham, UK, 2007.
- BULLEN, M. **A case study of participation and critical thinking in a university-level course delivered by computer conferencing**. Tese (Doutorado) — University of British Columbia, Vancouver, Canada, 2004.
- CAMPOS, F. C. A; SANTORO, F. M.; BORGES, M. R. S.; SANTOS, N. **Cooperação e aprendizagem on-line**. Rio de Janeiro: DP&A, 2003. (Col. Educação a distância). ISBN 85-7490-255-1.
- CANDOTTI, C. T. et al. Sistema hipermídia adaptativo baseado em estilos cognitivos. CINTED-UFRGS. **RENOTE - Revista Novas Tecnologias na Educação**. CINTED-UFRGS, [Porto Alegre], v. 4, n. 2. Dezembro, 2006. ISSN 1679-1916.
- CAPRETZ, L. F. Clues on Software Engineers' Learning Styles. **International Journal of Computing & Information Sciences**, v. 4, n.1. p. 46-49, April, 2006.

CAPRETZ, L. F. Implications of MBTI in Software Engineering Education. **ACM SIGCSE**, New York, v. 34 , n. 4, p. 134 – 137, December, 2002.

CAVELLUCCI, L. C. B. (2002). **Estilos de aprendizagem: um olhar para as diferenças individuais**. Universidade de Campinas, UNICAMP, 2002. Disponível em: http://www.iar.unicamp.br/disciplinas/am540_2003/lia/estilos_de_aprendizagem.pdf. Acesso em: 20 Fev 2012.

CHRISTAKIS, N.; FOWLER, J. **O poder das conexões: a importância do networking e como ele molda nossas vidas**. Rio de Janeiro: Campus-Elsevier, 2010. 308 p. ISBN 978-85-352-3047-5.

COLAS, F. **Data Mining Scenarios for the Discovery of Subtypes and the Comparison of Algorithms**. Tese (Doutorado) — Leiden Institute of Advanced Computer Science (LIACS), Faculty of Science, Leiden University, Leiden, The Netherlands, 2009.

COLAS, F.; BRAZDIL, P. Comparison of SVM and Some Older Classification Algorithms in Text Classification Tasks. In: BRAMER, M. (Ed.) **Artificial Intelligence in Theory and Practice**. Boston: Springer, 2006. v. 217, p. 169-178. (IFIP International Federation for Information Processing).

CORICH, S.; KINSHUK; HUNT, L. M. Computerised content analysis for measuring critical thinking within discussion forums. **Journal of e-Learning and Knowledge Society**, Roma, v. 2, n.1, p. 47-60, 2006.

CURTIS, D.; LAWSON, M. Exploring collaborative on-line learning. **Journal of Asynchronous Learning Networks**, vol. 5,1, p. 21-34, 2001. Disponível em: <http://sloanconsortium.org/jaln/v5n1/exploring-collaborative-online-learning>. Acesso em: 20 Fev 2012.

DAY, A. L.; CARROLL, S. Using an ability-based measure of emotional intelligence to predict individual performance, group performance, and group citizenship behaviours. **Personality and Individual Differences**, v.36, p. 1443-1458, 2004.

DE WEVER, B.; SCHELLENS T.; VALCKE, M.; VAN KEER H. Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. **Journal Computers and Education** - Methodological issue in researching CSCL, Oxford, UK, UK, v. 46, p. 6-28, January, 2006.

DESMEDT, E.; VALCKE, M. Mapping the learning styles ‘jungle’: an overview of the literature based on citation analysis. **Educational Psychology**, v. 24, n. 4, p. 445-464, 2004.

DIETTERICH, T. G. Ensemble Methods in Machine Learning. In: KITTLER, J.; ROLI, F. (Ed.). **INTERNATIONAL WORKSHOP ON MULTIPLE CLASSIFIER SYSTEMS – MCS**, 1., Jun, 2000, Cagliari, Italy. **Proceedings ...** London, UK.: Springer-Verlag, 21-23 June, 2000, p.1-15. (Lecture Notes In Computer Science, v. 1857). ISBN:3-540-67704-6.

DONMEZ, P.; ROSÉ, C. P.; STEGMANN, K.; WEINBERGER, A.; FISCHER, F. Supporting CSCL with Automatic Corpus Analysis Technology. In: TAK-WAI CHAN (Ed.). CONFERENCE ON COMPUTER SUPPORT FOR COLLABORATIVE LEARNING – CSCL'05. **Proceedings ...**, Taipei, Taiwan, May 30 - June 4, 2005: International Society of the Learning Sciences, 2005, p. 1-10. ISBN: 0-8058-5782-6.

DUARTE, J. C. **O Algoritmo Boosting at Start e suas Aplicações**. Tese (Doutorado) — Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2009.

DUNN, R.; De BELLO, T., BRENNAN, P., KRIMSKY, J.; MURRIN, P. Learning Style Researchers Define Differences Differently. **Educational Leadership**, February 1981. p. 372-375.

ELIA, M. F. O papel do Professor diante das Inovações Tecnológicas. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO – SBC 2008, 28., Workshop de Informática na Escola – WIE, 2008, Belém, Pará: **Anais eletrônicos...** Belém, Pará: UFPA, 2008. p. 215-224. Disponível em: <http://www.br-ie.org/pub/index.php/wie/article/view/980/966>. Acesso em: 20 Fev 2012.

ELIA, M. F.; SAMPAIO, F. F. Plataforma Interativa para Internet (PII): uma proposta de pesquisa-ação a distância para professores. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO – SBIE, 12., 2001, Vitória. **Anais eletrônicos...** Vitória: UFES, 2001. p. 102-109. Disponível em: <http://www.br-ie.org/pub/index.php/sbie/article/view/130/116>. Acesso em: 20 Fev 2012.

FELDER, R. M.; J. SPURLIN. Applications, Reliability, and Validity of the Index of Learning Styles. **International Journal of Engineering Education**, v. 21, n. 1, p. 103-112, 2005.

FELDER, R. M.; SILVERMAN, L. K. Learning Styles and Teaching Styles in Engineering Education. **Journal of Engineering Education**, v.78, n.7, p. 674-681, 1988.

FERREIRA, A. **Survey on Boosting Algorithms for Supervised and Semi-supervised Learning**. 2007. Relatório Técnico — Instituto Superior de Engenharia de Lisboa, 2007. Disponível em: <http://www.deetc.isel.ipl.pt/sistemastele/docentes/AF/Textos/RT/SurveyBoosting.pdf>. Acesso em: 20 Fev 2012.

FITZGERALD, N. et al. Exploiting conversational features to detect high-quality blog comments. In: BUTZ, C.; LINGRAS, P. (Eds.). Canadian Conference on Advances in Artificial Intelligence – AI'11, 24th, May, 25-27, 2011, St. John's, Canada. **Proceedings...** Berlin, Heidelberg: Springer-Verlag, 2011. p.122-127. ISBN: 978-3-642-21042-6.

FOGG, B. J. **Persuasive Technology**. Morgan Kaufmann, 2003. ISBN-10: 1558606432.

FORMAN, G.; KIRSHENBAUM, E.. Extremely Fast Text Feature Extraction for Classification and Indexing. In: ACM Conference on Information and Knowledge Management – CIKM '08, 17th, 26–30 October, 2008, Napa Valley California, USA.

Proceedings ... New York, NY, USA: Association for Computing Machinery, 2008. p. 1221-1230. ISBN: 978-1-59593-991-3. DOI: 10.1145/1458082.1458243.

FREIRE, P. **Pedagogia da Autonomia**. 43. ed. São Paulo: Paz e Terra, 2011. 144 p. ISBN: 9788577531639.

FURTER, M. J. K. **A relação entre os estilos de aprender e os estilos de ensinar em uma sala de aula multicultural de português língua estrangeira : um estudo de caso**. 2007. Dissertação (Mestrado em Lingüística) — Universidade de Brasília, Brasília, 2007.

GAMA, J. M. P. **Combining Classification Algorithms**. 1999. 195 f. Tese (Doutorado em Ciências da Computação) — Faculdade de Ciências, Universidade do Porto, Portugal, 1999. Disponível em: <http://repositorio-aberto.up.pt/handle/10216/10017>. Acesso em: 20 Fev 2012.

GARCIA, V. et al. The class imbalance problem in pattern classification and learning. In: Congreso Español de Informática, 2, 2007. Zaragoza, Spain. **Taller de Minería de Datos y Aprendizaje – TAMIDA 2007**, 4. Zaragoza, Spain, September, 2007. ISBN: 978-84-9732-602-5.

GELLER, M.; TAROUÇO, L. M. R.; FRANCO, S. R. K. Educação a Distância e Estilos Cognitivos: Construindo a Adaptação de Ambientes Virtuais. In: Congresso Iberoamericano de Informática Educativa, 7., 2004, Monterrey - México. **VII Congresso Iberoamericano de Informática Educativa**, 2004. v. II. Disponível em: <http://www.niee.ufrgs.br/eventos/RIBIE/2004/comunicacao/com274-283.pdf>. Acesso em: 20 Fev 2012.

GILLIES, R. M. Structuring cooperative group work in classrooms. In: **International Journal of Educational Research**, v. 39, pp. 35–49, 2003.

GIORGINO, T. **An Introduction to Text Classification**. Tese (Doutorado) — The Scuola Avanzata di Formazione Integrata, Istituto Universitario di Studi Superiori di Pavia, 2004.

GLADWELL, M. **O Ponto da Virada**. Rio de Janeiro: Sextante, 2009. ISBN: 9788575424834.

GLENN, J. M. Teaching the Net Generation. **Business Education Forum**, 2000. v. 54, n.3, p. 6-14.

GOMES NETO, G.; GOMES, A. S.; TEDESCO, P. A. Elicitação de Requisitos de Sistemas Colaborativos de Aprendizagem Centrada na Atividade de Grupo. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO – SBIE, 14., 2003, Rio de Janeiro. **Anais eletrônicos...** Rio de Janeiro: UFRJ, 2003. Disponível em: <http://www.br-ie.org/pub/index.php/sbie/article/view/262/248>. Acesso em: 20 Fev 2012.

GRAF, S.; KINSHUK. Advanced Adaptivity in Learning Management Systems by Considering Learning Styles. In: IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology – WI-IAT '09, 2009. **Proceedings ...**

Washington, DC, USA: IEEE Computer Society, 2009. v 03, p. 235-238. ISBN: 978-0-7695-3801-3. DOI: 10.1109/WI-IAT.2009.271

GRAF, S.; LAN, C. H.; LIU, T. C.; KINSHUK. Investigations about the Effects and Effectiveness of Adaptivity for Students with different Learning Styles. In: IEEE International Conference on Advanced Learning Technologies – ICALT'09, 9th, 15-17 July 2009, 2009, Riga, Latvia. **Proceedings ...** p. 415-419.

GREGORY, J. M. **As 7 Leis do Ensino**. Rio de Janeiro: JUERP, 1993.

GUILD, P. B.; GARGER, S. **Marching to Different Drummers**. ACSD, 1985.

GUNAWARDENA, C. N.; CARABAJAL, K.; LOWE, C. A. Critical analysis of models and methods used to evaluate online learning networks. In: ANNUAL MEETING OF THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, 2001, Seattle. **Reports...**Seattle: American Educational Research Association, 2001.

GUNAWARDENA, C. N.; LOWE, C. A.; ANDERSON, T. Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. **Journal of Educational Computing Research**, Amityville, v, 17, n. 4, p. 397-431, 1997.

HAKKARAINEN, K. P. J. **Epistemology of scientific inquiry and computer-supported collaborative learning**. 1998. 347 f. Tese (Doutorado) — University of Toronto, Toronto, Ont., Canada, 1998. ISBN:0-612-41435-3.

HALE, M.; JORGENSEN, N.; GAMBLE, R. Predicting individual performance in student project teams. In: **IEEE-CS Conference on Software Engineering Education and Training – CSEE&T**, 24th, 2011, Honolulu, HI USA. p.11-20, 22-24 May 2011. ISBN: 978-1-4577-0349-2.

HAY, L. E. (2000). Educating the Net Generation. In: **School Administrator**, v. 57, n. 54, p. 6-10, April 2000. Disponível em: <http://www.aasa.org/SchoolAdministratorArticle.aspx?id=14422>. Acesso em: 20 Fev 2012.

HENDRICKS, H. Ensinando para transformar vidas. Rio de Janeiro : BETÂNIA, 1991.

HENRI, F. Collaborative learning through computer conferencing. In: KAYE, A.R. (Ed.). NATO Advanced Research Workshop on Collaborative Learning and Computer Conferencing, July 29-August 3, 1991, Copenhagen, Denmark. **Proceedings...**, Berlin: Springer-Verlag, 1992. NATO ASI series. Series F, Computer and system sciences, vol. 90, NATO. ISBN 10 3540557555, 0387557555.

HIROAKI, Hashiura et al. A software development group exercise support environment, EtUDE: the system overview and the system evaluation through applying to classes. In: KAZOVSKY, L.; MASTORAKIS, N.; KURI-MORALES, A.; SAKELLARI, I. (Eds.). WSEAS International Conference on Software Engineering, Parallel and Distributed Systems

– SEPADS'08, 7th, 2008, Cambridge, UK. **Proceedings of...** Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2008. p. 124-131. (Electrical and Computer Engineering Series). ISBN: 978-960-6766-42-8.

HIROAKI, Hashiura et al. A system for supporting group exercise in software development with facilities to create an optimal plan of student grouping and team formation of each group. In: KAZOVSKY, L.; MASTORAKIS, N.; KURI-MORALES, A.; SAKELLARI, I. (Eds.). WSEAS International Conference on Software Engineering, Parallel and Distributed Systems – SEPADS'08, 7th, 2008, Cambridge, UK. **Proceedings of...** Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2008. p. 149-157. (Electrical and Computer Engineering Series). ISBN: 978-960-6766-42-8.

HOWARD, K. C. **Millennials spur teaching change**. Las Vegas Review Journal. 6 March 2006. disponível em: http://www.reviewjournal.com/lvrj_home/2006/Mar-06-Mon-2006/news/6146322.html. Acesso em: 20 Fev 2012.

HUMMEL, H. G. K. et al. Fostering Participation in Learning Networks by Using Reward Systems and Face-to-Face Meetings. In: **International Conference on Advanced Learning Technologies – ICALT 2006**, 6th, 5-7 July 2006. p. 534-536.

HWANJO Yu; CHENGXIANG Zhai; JIAWEI Han. Text classification from positive and unlabeled documents. In: International Conference on Information and Knowledge Management – CIKM '03, 20th, November 03-08, 2003, New Orleans, LA, US. **Proceedings ...** New York, NY, US: ACM, 2003. P. 232–239. ISBN: 1-58113-723-0. DOI: 10.1145/956863.956909.

IKONOMAKIS, M.; KOTSIANTIS, S.; TAMPAKAS, V. Text Classification Using Machine Learning Techniques. **Wseas Transactions On Computers**, v. 4, n. 8, p. 966–974. August, 2005. Disponível em: http://www.infoautoclassification.org/public/articles/Ikonomakis-et.-al._Text-Classification-Using-Machine-Learning-Techniques.pdf. Acesso em: 20 Fev 2012.

JINGNIAN, Chen et al. Feature selection for text classification with Naïve Bayes. **Expert Systems with Applications: An International Journal**, Tarrytown, NY, USA, v.36, n.3, p. 5432-5435, April, 2009.

KAO, A.; POTEET, R. S. **Natural Language Processing and Text Mining**. Londres: Springer-Verlag. 2007. 277 p. ISBN-13: 978-1846281754.

KEYSERS, C. Mirror Neurons. **Current Biology**, v. 19, n. 21, p. 971–973. DOI:10.1016/j.cub.2009.08.026. PMID 19922849.

KIAN, Ming Adam Chai; HAI, Leong Chieu; HWEE, Tou Ng. Bayesian online classifiers for text classification and filtering. In: 25th annual international ACM SIGIR conference on Research and development in information retrieval, August 11-15, 2002, Tampere, Finland. **Proceedings...** New York, NY: ACM, 2002. p. 97-104. ISBN: 1-58113-561-0.

KING, A.; ROSENSHINE, B. Effects of guided cooperative questioning on children's knowledge construction. **Journal of Experimental Education**, v. 61, n. 2, p. 127–148, Winter, 1993.

KINSHUK; GRAF, S. Considering Cognitive Traits and Learning Styles to Open Web-Based Learning to a Larger Student Community. In: JEMNI, (Ed.). International Conference on Information and Communication Technology and Accessibility – ICTA 2007, April 2007, Hammamet, Tunisia. **Proceedings...** ICTA, p. 21-26. Disponível em: www.esstt.rnu.tn/utic/tica2007/sys_files/medias/docs/p04.pdf. Acesso em: 20 Fev 2012.

KOLB, D. A. **Experiential learning: Experience as the source of learning and development**. New Jersey: Prentice-Hall. 1984.

KOTSIANTIS, S.; KANELLOPOULOS, D.; PINTELAS, P. Handling imbalanced datasets: A review. In **GESTS International Transactions on Computer Science and Engineering**. v.30, n.1, p. 25-36, 2006. Disponível em: <http://www.math.upatras.gr/~esdlab/en/members/kotsiantis/imbalanced%20datasets%20survey%20paper%20gests.pdf>. Acesso em: 20 Fev 2012.

KOTSIANTIS, S.; PINTELAS, P. Mixture of Expert Agents for Handling Imbalanced Data Sets. **Annals of Mathematics, Computing & TeleInformatics**, v.1, n.1, p.46-55, 2003.

KRAEMER, Maria Elisabeth Pereira. A Avaliação da Aprendizagem como Processo Construtivo de um Novo Fazer. **Avaliação**, Campinas, v. 10, n. 2, p. 137-147, 2005.

KRIPPENDORFF, K. **Content analysis: an introduction to its methodology**. 2^a ed. Newbury Park: Sage Publications, Inc, December, 2003.

KULJIS, J.; LINES, L. (2007). The role of web-based learning environments in fostering collaboration. In: CONSTANTINE STEPHANIDIS (Ed.). INTERNATIONAL CONFERENCE ON UNIVERSAL ACCESS IN HUMAN-COMPUTER INTERACTION – UAHCI'07, 4th, 2007. **Proceedings ...** Berlin, Heidelberg: Springer-Verlag, 2007. p. 667-676.

KUNCHEVA, L.I. **Combining Pattern Classifier: Methods and Algorithms**. John Wiley & Sons, 2004.

LAYMAN, L.; CORNWELL; T.; WILLIAMS, L. Personality types, learning styles, and an agile approach to software engineering education. In: Technical Symposium on Computer Science Education – SIGCSE '06, 37th, March 1-5, 2006, Houston, Texas USA. **Proceedings ...** New York, NY: ACM, 2006, p. 428–432. ISBN:1-59593-259-3. DOI:10.1145/1121341.1121474.

LAZONDER, A. W.; WILHELM, P.; OOTES, S. A. W. Using sentence openers to foster student interaction in computer-mediated learning environments. In: **Computers & Education**, 2003. v. 41, p. 291–308.

LEWIS, D. D.; RINGUETTE, M. A comparison of two learning algorithms for text categorization. In: Annual Symposium on Document Analysis and Information Retrieval, 3th, 1994. p. 81-93. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.860>. Acesso em: 20 Fev 2012.

LIJUAN Jiao; LIPING Feng. Text Classification Based on Ant Colony Optimization. International Conference on Information and Computing – ICIC '10, 3rd, 2010, Wuxi, China. **Proceedings...** Washington, DC, USA: IEEE Computer Society, 2010. v.3, p. 220 – 232. ISBN: 978-0-7695-4047-4.

LOPES, M. S. S. **Avaliação da Aprendizagem em Atividades Colaborativas em EAD Viabilizada por um Forum Categorizado**. Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2007.

LUI, A. K.; LI, S. C.; CHOY, S. O. An Evaluation of Automatic Text Categorization in Online Discussion Analysis. In: **International Conference on Advanced Learning Technologies – CALT 2007**, 7th, 18-20 July 2007, p. 205 - 209.

LUI, A.K. A Learning Thermometer: Improving Visibility of Learning Activities in Online Discussion Forums. **Distance Education and Technology: Issues and Practice**, Open University of Hong Kong Press, p. 293-307, 2004.

MACDONALD, J. Assessing online collaborative learning: process and product. **Computers & Education**, v. 40, p. 377–391, 2003.

MARTINEZ, M.; BUNDERSON, C. V. **Foundations for Personalized Web Learning Environments**. Edumetrics Institute and Brigham Young University. Provo, UT. 2008.

MCLAREN, B. M et al. Using Machine Learning Techniques to Analyze and Support Mediation of Student E-Discussions. In: LUCKIN, R.; KOEDINGER, K. R.; GREER, J. (Eds.). International Conference on Artificial Intelligence in Education – AIED 2007, 13th, 2007. **Proceedings ...** The Netherlands: IOS Press Amsterdam, 2007. p. 331-338. ISBN: 978-1-58603-764-2.

MENESES, P. **Hegel e a Fenomenologia do Espírito**. Rio de Janeiro: Zahar, 2003. (Coleção Filosofia Passo-a-Passo).

MEYER, K. Evaluating online discussions: four different frames of analysis. **Journal of Asynchronous Learning Networks**, v.8, n.2, p. 101–114, 2004.

MING Lai. The Role of Argumentation in Online Knowledge Building Activities. In: International Conference on Computers in Education – ICCE 2008, 16th, October, 27-31, 2008, Taipei, Taiwan. **Proceedings ...** Taipei, Taiwan: National Central University, Graduate Institute of Network Learning Technology, 2008. p. 317–324. Disponível em: http://www.apsce.net/icce2008/contents/proceeding_0317.pdf. Acessado em 20 Fev 2012.

MOLDOVAN, D.; SURDEANU, M. (2002). On the role of information retrieval and information extraction in question answering systems. In: PAZIENZA, M. T. (Ed.). **Information Extraction in the Web Era**. Julho 2002. (Lecture Notes in Computer Science, 2003, v. 2700, , p. 129-147). DOI: 10.1007/978-3-540-45092-4_6.

MONTANES,E et al.. Measures of Rule Quality for Feature Selection in Text Categorization In: **Iinternational Symposium on Intelligent Data Analysis**, 5th, 2003, Germeny: Springer-Verlag, 2003, v. 2810, p. 589-598.

MOREIRA, M. A.. **Aprendizagem significativa**. Brasília: Editora Universidade de Brasília, 1999.

MORIN, E. **Os Sete Saberes necessários à Educação do Futuro**. 10. ed. Brasília, DF: UNESCO, 2005.

NEWMAN, D. R.; WEBB, B.; COCHRANE, C. A content analysis method to measure critical thinking in face-toface and computer supported group learning. **Interpersonal Computing and Technology**, v.3, p. 56–77, April 1995.

NIGAM, K., LAFFERTY, J., McCALLUM, A. Using maximum entropy for text classification. In: **Workshop on Machine Learning for Information Filtering – IJCAI’99**, p. 61-67, 1999.

NLTK. Disponível em: <http://nltk.sourceforge.net/index.php/Book>. Acessado em 20 Fev 2012.

NUNES, L. C.; VILARINHO, L. R. G. Avaliação da aprendizagem no ensino online: em busca de novas práticas. In: SILVA, M.; SANTOS, E. (Org.). **Avaliação da aprendizagem em educação online**. São Paulo: Loyola, 2006. p. 109-122..

OLIVEIRA, M. K. **Vygotsky: aprendizado e desenvolvimento: um processo sócio-histórico**. São Paulo: Scipione, 1997. 111p.

OSER, K. 2005. Kids cram more hours in media day. **Advertising Age**. 2005. v. 76, n. 46, p. 31.

PAPANIKOLAOU, K. A.; GRIGORIADOU, M. **Accomodating learning style characteristics in Adaptive Educational Hypermedia Systems**. Department of Informatics & Telecommunications, University of Athens, Panepistimiopolis, GR- 15784, Athens, Greece, 2004.

PAPERT, S. **The children’s machine: Rethinking school in the age of the computer**. New York: Basic Books, 1993.

PAREDES, P.; RODRIGUEZ, P. The Application of Learning Styles in Both Individual and Collaborative Learning. In: **International Conference on Advanced Learning Technologies – ICALT 2006**, 6th, 5-7 July 2006. p. 1141-1142.

PARVEZ, S. M.; BLANK, G. D. Individualizing Tutoring with Learning Style Based Feedback. **Lecture Notes in Computer Science**, 5091/2008, p. 291-301.

PEDROSA DE JESUS, H. P.; MOREIRA, A. C. The Role of Students' Questions in Aligning Teaching, Learning and Assessment: A Case Study from Undergraduate Sciences. **Assessment & Evaluation in Higher Education**, v. 34, n. 2 p. 193-208, April 2009.

PENA-SHAFF, J. B.; NICHOLLS, C. Analyzing student interactions and meaning construction in computer bulletin board discussions. **Computers & Education**, v. 42, p. 243–265, 2004.

POPESCU, E. Evaluating the Impact of Adaptation to Learning Styles in a Web-Based Educational System. International Conference on Advances in Web Based Learning – ICWL 2009, 8th, 2009. **Proceedings ...** Berlin, Heidelberg: Springer-Verlag, 2009. (Lecture Notes in Computer Science, v.5686, p.343-352). ISBN 978-3-642-03425-1.

POPESCU, E.; BADICA, C. Providing Personalized Courses in a Web-Supported Learning Environment. In: International Joint Conference on Web Intelligence and Intelligent Agent Technology – WI-IAT'09 (workshop SPeL), 2009. **Proceedings ...** Washington, DC, USA: IEEE Computer Society, 2009. v. 03, p. 239-242,. ISBN: 978-0-7695-3801-3.DOI: 10.1109/WI-IAT.2009.272

POZO, J.I. Estrategias de aprendizaje. **Educar**. v.1, n.3, 1998.

PRENSKY, M. **Don't bother me Mom-I'm learning**. Minneapolis: Paragon House, 2006.

RAGAS, H.; KOSTER, C. H. Four text classification algorithms compared on a Dutch corpus. In: Research and Development in Information Retrieval – SIGIR '98, 21st, 1998. **Proceedings ...** New York, NY, USA: ACM, 1998. p. 369 - 370. ISBN:1-58113-015-5. DOI: 10.1145/290941.291059.

RATNAPARKHI, A. A Simple Introduction to Maximum Entropy Models for Natural Language Processing. **IRCS Report**, August 97, University of Pennsylvania, Philadelphia, PA, 1997. DOI: 10.1.1.43.3081. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.43.3081>. Acesso em: 20 Fev 12

RICHETTI, C.; SHEERIN, J. Helping students ask the right questions. **Educational Leadership**, v. 57, n.3, p. 58-62, 1999.

ROGERS, E. M. Examinations: Powerful Agents for Good or Ill in Teaching. **American Journal of Physics**, 30, 10, p. 954-962, 1960B.

ROGERS, E. M. **Physics for the inquiring mind: The methods, nature & philosophy of physical science** Princeton University Press, 1960A.

ROQUE, G. O. B. Uma **proposta de um modelo de avaliação de aprendizagem por competências para cursos a distância baseados na web**. Dissertação (Mestrado em

Informática) Universidade Federal do Rio de Janeiro. Instituto de Matemática. Núcleo de Computação Eletrônica, 2004.

ROSÉ, C. et al. Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. **Computer-Supported Collaborative Learning**, v.3, p. 237–271, 2008. DOI 10.1007/s11412-007-9034-0.

ROURKE, L. Assessing social presence in asynchronous, text-based computer conferencing. **Journal of Distance Education**, v. 14, n.2, 1999.

ROURKE, L., et al. Methodological issues in the content analysis of computer conference transcripts. **International Journal of Artificial Intelligence in Education**, v. 12, 2001.

RUIZ, M. P. P. et al. Adaptation in current e-learning systems. **Computer Standards & Interfaces**, v.30, n.1-2, p.62-70, January, 2008.

RUSSO, J. E.; SCHOEMAKER, P. J. H. **Decisões Vencedoras**. Rio de Janeiro: Campus, 2002. 390 p. ISBN 85-352-0974-3.

SAEED, N.; YANG, Y.; SINNAPPAN. Emerging Web Technologies in Higher Education: A Case of Incorporating Blogs, Podcasts and Social Bookmarks in a Web Programming Course based on Student's Learning Styles and Technology Preferences. **Educational Technology & Society**, v. 12, n. 4, p. 98-109, 2009.

SANG-BUM Kim et al. Effective Methods for Improving Naïve Bayes Text Classification. In: Pacific Rim International Conference on Artificial Intelligence – PRICAI 2002, 7th, August 18-22, 2002, Tokyo, Japan. **Proceedings ...** Berlin: Springer, 2002, p. 479–484. (Lecture Notes in Computer Science, v. 2417/2002). ISBN: 3-540-44038-0.

SANTOS, A. A. A.; BARIANI, I. C.; CERQUEIRA, T. C. Estilos cognitivos e estilos de aprendizagem. **Leituras de psicologia para formação de professores**. 2. ed. Petrópolis, RJ: Vozes, 2000, p. 44-57.

SANTOS, E. M. **Avaliação da influência dos estilos cognitivos no perfil do aluno de educação a distância**. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2007.

SCHEUER, O.; MCLAREN, B. M. Helping Teachers Handle the Flood of Data in Online Student Discussions. In: WOOLF, B. et al. (Eds.). International Conference on Intelligent Tutoring Systems – ITS-08, 9th, 23-27 June, 2008, Montreal, Canada. **Proceedings ...** Berlin: Springer, 2008. (Lecture Notes in Computer Science, v. 5091, p. 323-332).

SCHÜTZE, H.; HULL D. A.; PEDERSEN, J. O. A comparison of classifiers and document representations for the routing problem. In: FOX, E. A.; INGWERSEN, P.; FIDEL, R. (Eds.). International Conference on Research and Development in Information Retrieval –

SIGIR, 18th, July 09-13, 1995, Seattle, Washington, US. **Proceedings ...** ACM Press, 1995. p. 229-237. ISBN: 0-89791-714-6.

SEBASTIANI, F. Machine learning in automated text categorization. **Journal ACM Computing Surveys – CSUR**, New York, NY, USA, v. 34, p.1 – 47, March 2002. ISSN: 0360-0300 EISSN: 1557-7341 doi>10.1145/505282.505283.

SERCE, F. et al. Exploring the communication behavior among global software development learners. **International Journal of Computer Applications in Technology**, v. 40, n.3, p. 203 - 215, 2011.

SHI Yong-feng; ZHAO Yan-ping. Comparison of text categorization algorithms. **Wuhan University Journal of Nature Sciences**, v. 9, n. 5, p. 798-804, May 2004. DOI: 10.1007/BF02831684.

SIEMENS, G. Connectivism: A learning theory for the digital age. **International. Journal of Instructional Technology and Distance Learning**, v.2, p.10, 2005.

SOLLER, A. et al. Promoting effective peer interaction in an intelligent collaborative learning system. In: GOETTL, B. P. et al. (Eds.). International Conference on Intelligent Tutoring Systems, 4th, 1999, San Antonio, TX, US. **Proceedings ...** London, UK: Springer-Verlag, 1998. pp. 186-195. (Lecture Notes In Computer Science). ISBN: 3-540-64770-8.

SOLLER, A.; LESGOLD, A. Modeling the Process of Collaborative Learning. In: International Workshop on New Technologies in Collaborative Learning, 2000, Awaiji-Yumebutai, Japan. **Proceedings ...** Awaiji-Yumebutai, Japan, 2000.

SONG, M. H. et al. Automatic Classification of Web pages based on the Concept of Domain Ontology. In: Asia-Pacific Software Engineering Conference – APSEC '05, 12th, 15-17 Dec. 2005, Taipei, Taiwan. **Proceedings ...** Taipei, Taiwan, 2005.

SU Kim et al. Automatic classification of sentences to support Evidence Based Medicine. BMC bioinformatics, 12. Suppl 2. 2011. In: DOHEON, L.; HAGIT, S.; SHAMKANT, B. N.; MIN Song (Eds.). International Workshop on Data and Text Mining in Biomedical Informatics – DTMBio 2010, 4th, 26 October 2010, Toronto, Canada. **Proceedings...** BMC Bioinformatics 2011, 12(Suppl 2):S5, 29 March 2011. doi:10.1186/1471-2105-12-S2-S5. Disponível em: <http://www.biomedcentral.com/1471-2105/12/S2/S5>. Acesso em: 20 Fev 2012.

SU Nam Kim; CAVEDON; L.; BALDWIN, T. Classifying dialogue acts in one-on-one live chats. In: Conference on Empirical Methods in Natural Language Processing – EMNLP '10, October 09-11, 2010, Massachusetts, USA. **Proceedings...** Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. p.862-871.

SUROWIECKI, J. **The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations**. Little Brown. ISBN 0-316-86173-1. 2004.

SWIGGER, K.; et al. Using Content and Text Classification Methods to Characterize Team Performance. In: IEEE International Conference on Global Software Engineering – ICGSE, 5th, 23-26 August, 2010, Princeton, New Jersey, US. **Proceedings...** Princeton, New Jersey, US, 2010. p.192-200. ISBN: 978-0-7695-4122-8.

TAPSCOTT, D. **A hora da geração digital**. Rio de Janeiro: Agir, 2010.

TATTERSALL, C. et al. Swarm-Based Adaptation: Wayfinding Support for Lifelong Learners. In: DE BRA, P.; NEJDL, W. (Eds.). International Conference Adaptive Hypermedia and Adaptive Web-Based Systems – AH 2004, 3th, August 23-26, 2004, Eindhoven, The Netherlands. **Proceedings...** Springer, 2004. (Lecture Notes in Computer Science, v.3137, p. 336-339). ISBN 3-540-22895-0.

TENENBAUM, L.; SHAPIRA, B.; SHOVAL, Peretz (2008). “Ontology-Based Classification Of News In An Electronic Newspaper. In: International Conference Intelligent Information and Engineering Systems – INFOS 2008, June-July 2008, Varna, Bulgaria. **Proceedings ...** Varna, Bulgaria, 2008.

THE MCGRAW HILL SMALL GROUP COMMUNICATIONS. 2011. Disponível em: http://www.mhhe.com/socscience/comm/group/stu_index.html. Acesso em: 20 Fev 2012.

TOLMIE, A. et al. The effects of adult guidance and peer discussion on the development of children’s representations: Evidence from the training of pedestrian skills. **British Journal of Educational Psychology**, v. 96, p. 181–204, 2005.

VAN DER POL, J. Identifying and modeling variables in complex CSCL-situations. Case study: The use of asynchronous electronic discussions. In: **Workshop on Designing Computational Models of Collaborative Learning Interaction – CSCL2002**, January, 2002, Boulder, CO.

VANLEHN, K et al. What makes a tutorial event effective? In: GERNSBACHER, M. A.; DERRY, S. (Eds.). **Conference of the Cognitive Science Society**, 20th, 1998, Hillsdale, NJ: Erlbaum, 1998. p. 1084-1089.

WEBB, N. M. The teacher’s role in promoting collaborative dialogue in the classroom. **British Journal of Educational Psychology**, v. 79, p. 1–28, 2009.

WEISS, S. M et al. **Text Mining: Predictive methods for analyzing unstructured information**. USA: Springer. 2005. p 66-69.

WIEBE, J. et al. Learning Subjective Language. **Computational Linguistics**, v. 30, n. 3, p.277–308, 2004.

XU-YING Liu; JIANXIN Wu ; ZHI-HUA Zhou. Exploratory Under-Sampling for Class-Imbalance Learning. In: **IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics**, v. 39 , n. 2, p. 539 - 550, April 2009.

YEH, A. S.; HIRSCHMAN, L.; MORGAN, A. A. Evaluation of text data mining for database curation: lessons learned from the KDD challenge cup. **Bioinformatics**, v.19, (supp.1), p. i331-i339, 2003.

YI-HSING, Chang; HSIU-YI, Huang. An Automatic Document Classifier System Based On Naïve Bayes Classifier And Ontology. In: **Seventh International Conference on Machine Learning and Cybernetics**, 12-15 July, 2008, Kunming, v. 6, p. 3144 - 3149. ISBN: 978-1-4244-2095-7.

YIMING Yang; XIN Liu. A re-examination of text categorization methods. In: International ACM Conference on Research and Development in Information Retrieval – SIGIR '99, August 15-19, 1999, Berkeley, California, US. **Proceedings ...** New York, NY, US: ACM, 1999. p.42-49. ISBN: 1-58113-096-1. DOI: 10.1145/312624.312647.

ZHIPENG Xie. Effective Boosting of Naïve Bayesian Classifiers by Local Accuracy Estimation. In: THEERAMUNKONG, T. et al. (Eds.). Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining – PAKDD '09, 13th, April 27-30, 2009, Bangkok, Thailand. **Proceedings ...** Berlin, Heidelberg: Springer-Verlag, 2009. p. 849-858. ISBN: 978-3-642-01306-5. DOI: 10.1007/978-3-642-01307-2_88.

ANEXO 1

Quadro 5. Fases de aprendizagem, segundo Gunawardena, Lowe e Anderson (1997).
Fonte: (LOPES, 2007)

FASE	CARACTERÍSTICA	DETALHAMENTO
1	Compartilhamento/comparação de informações	Fase inicial, na qual há compartilhamento, comparação de informações e percepções, representando o estágio no qual os participantes apresentam afirmativas sobre seus entendimentos relacionados ao assunto em discussão e esclarecem detalhes.
2	Descoberta e exploração de discordâncias	É a presença de desarmonia e inconsistência entre as idéias expressadas, que exige exploração de pontos divergentes.
3	Negociação de significado e co-construção do conhecimento	Trata-se da fase de negociação de conflitos, na qual são propostas co-construções para integração e consenso.
4	Teste e modificação de sínteses propostas/co-construção	Fase em que as novas idéias e co-construções são testadas e comparadas com posições individuais e outras fontes.
5	Entendimento/aplicação de novas co-construções	A última fase apresenta idéias com a versão final da construção do grupo.

ANEXO 2

Quadro 6 - Categorias propostas por Lopes (2007)

Categoria	Subcategoria	Significado
Social	Agradecimento	Sentenças que expressam reconhecimento e gratidão e não estão diretamente relacionadas à solução de problema ou ao desenvolvimento de raciocínio e aprendizagem.
	Saudação	Cumprimento, anunciando a chegada ou a saída do ambiente de discussão.
Confirmação	Co-construção	Sentenças declaradas pelo debatedor após o conhecimento de uma afirmação ou informação de outro debatedor, que confirmam uma posição a favor ou contrária ao que foi declarado.
Esclarecimento	Definição	Frases que exprimem a intenção de identificar ou construir um conceito a fim de esclarecer uma dúvida trazida ao debate.
	Dúvida	Sentenças nas quais são expressas incertezas sobre o assunto debatido ou algo apresentado em referências.
	Esclarecimento	Sentenças nas quais o debatedor esclarece um conceito ou uma dúvida apresentada no debate.
	Negociação	Frases nas quais o debatedor chama os outros debatedores a estabelecer consenso diante de algum ponto de dúvida.
	Opinião	Sentenças nas quais é expressa a motivação de se ter tomado uma decisão ou estabelecido um conceito ou uma posição.
Informação	Compartilhamento	Expressões nas quais uma informação é participada aos demais componentes do debate, a fim de que possam tomar um posicionamento sobre o assunto.
	Confirmação	Sentenças nas quais é retomado um assunto já tratado a fim de apoiar uma opinião, confirmar uma regra ou para concretizar a verdade de uma afirmação.
	Definição	Expressões que explicam a significação de um termo.
	Opinião	Frases que expõem um juízo ou uma convicção do debatedor sobre determinado assunto.
	Posição	Expressões que trazem ao debate uma idéia, um parecer sobre algo já exposto.
	Validação-Síntese	Sentenças que relatam de uma forma única a posição do grupo sobre algum assunto discutido após o estabelecimento de um consenso.
Mediação	Mediação	Sentenças que exprimem uma intervenção que vise facilitar ou desenvolver a participação do grupo.
Motivação	Motivação	Frases que representem um estímulo à participação do grupo ou de um outro participante.
Negociação	Co-construção	Sentenças nas quais o debatedor solicita um posicionamento dos outros debatedores sobre algo que foi apresentado na discussão.
	Concordância	Expressões nas quais o debatedor exprime a idéia de estar de acordo com algo que foi apresentado no debate.

Negociação	Discordância	Expressões declarando que o debatedor não está de acordo com algo que foi declarado no debate.
	Negociação	Sentenças que representam a posição de concordância ou discordância sobre a opinião de outro participante, a busca do posicionamento de outro participante ou o questionamento sobre uma posição de outro participante.
Tarefa	Finalização	Frases que anunciam a intenção de algum debatedor encerrar a discussão para a resolução da tarefa.
	Inicialização	Frases que anunciam a intenção de algum debatedor iniciar a discussão para a resolução da tarefa.
	Sumarização	Frases que expressam uma proposta de desfecho ou conclusão de uma idéia elaborada pelo grupo.
Texto	Co-construção	Sentenças que apresentam uma idéia do grupo ou de um debatedor para complementar um texto de referência.
	Definição	Frases que se reportam a uma posição de um autor de alguma referência apresentada no enunciado da tarefa ou no decorrer do debate.
	Discordância	Frases nas quais o debatedor evidencia que não concorda com algo que foi apresentado num texto de referência.
	Esclarecimento	Frases nas quais o debatedor solicita a elucidação de algum trecho de um texto de referência.
	Modificação	Expressões pelas quais o debatedor apresenta uma outra referência que complementa o debate ou um texto apresentado na referência.

ANEXO 3

Quadro 7- Relação entre as categorias de Lopes (2007) e as fases de Gunawardena, Lowe e Anderson

Categoria	Subcategoria	Início de Sentença	Fase
Social	Agradecimento	Obrigado	0
Social	Saudação	Olá	0
Social	Saudação	Até logo	0
Confirmação	Co-construção	Ok.	1
Confirmação	Afirmação	Sim.	1
Confirmação	Negação	Não.	3
Confirmação	Co-construção	Exatamente.	1
Esclarecimento	Definição	O que significa	1
Esclarecimento	Definição	Existe diferença entre	1
Esclarecimento	Dúvida	Entendi	1
Esclarecimento	Dúvida	Alguém lembra	1
Esclarecimento	Dúvida	Não sei	1
Esclarecimento	Dúvida	Não entendi	1
Esclarecimento	Dúvida	Talvez	1
Esclarecimento	Dúvida	Por que	1
Esclarecimento	Dúvida	Você entendeu	1
Esclarecimento	Dúvida	Não me lembro	1
Esclarecimento	Esclarecimento	Eu entendo que	2
Esclarecimento	Negociação	Vocês não acham que	2
Esclarecimento	Negociação	Não teríamos que	3
Esclarecimento	Negociação	Como nós vamos	3
Esclarecimento	Opinião	Por esta razão	3
Informação	Compartilhamento	Gostaria de compartilhar	1
Informação	Confirmação	Por exemplo	2
Informação	Definição	O termo significa	1
Informação	Opinião	Eu acho que	1
Informação	Posição	Eu sugiro que	3
Informação	Posição	A idéia é interessante	1
Informação	Posição	Partindo do princípio	2
Informação	Validação- síntese	Poderíamos afirmar que	4
Informação	Validação-síntese	Podemos definir que	4
Mediação	Intervenção	Vamos participar	0
Motivação	Motivação	Excelente!	2
Motivação	Motivação	Muito bom	0
Negociação	Co-construção	Vocês concordam	2
Negociação	Concordância	Concordo	2
Negociação	Discordância	Não seria o caso	2
Negociação	Discordância	Discordo	2
Negociação	Discordância	Entendi de forma diferente	3
Negociação	Negociação	Porém	3
Negociação	Negociação	Questiono	3
Tarefa	Inicialização	Vamos iniciar	4
Tarefa	Finalização	Vamos terminar	4
Tarefa	Sumarização	A visão final do grupo	5
Tarefa	Sumarização	Podemos concluir	5
Texto	Co-construção	Em acréscimo ao que é dito no texto	3
Texto	Definição	O autor apresenta no texto	1
Texto	Discordância	Não há no texto	3
Texto	Discordância	Discordo do texto	3
Texto	Esclarecimento	Vamos rever	2
Texto	Complementação	Outro autor	4