

Uma Proposta de Avaliação Automática de Mapas Conceituais para Ambientes de Ensino a Distância

Vanessa Martins Caldas¹, Eloi Luiz Favero¹

¹Centro de Ciências Exatas e Naturais - Universidade Federal do Pará (UFPA)
Belém-PA-Brasil
{vanessam, eloi.favero}@ufpa.br

Abstract. Automated Evaluation is a required feature in Virtual Learning Environment. Nowadays, this kind of evaluation is reality only for closed questions (objective questions). Conceptual Maps is learning and evaluation methodology that handle opened question, whose methods to do automated evaluation has been researched. This paper describes an automated evaluation proposal for Conceptual Maps in a Virtual Learning Environment context. The proposal is centered in Artificial Intelligence: classification techniques (N-Grams and KNN). Results achieve an average of 90% rights related to human experts.

Key Words: Automated Evaluation, Virtual Learning Environment, Conceptual Maps, Artificial Intelligence, N-Grams, KNN.

1. Introdução

A avaliação discente é uma tarefa que possui como propósito principal verificar a aprendizagem de um aluno acerca de um assunto. A avaliação do conhecimento fundamenta-se na aplicação de testes e exercícios e na correção dos mesmos atribuindo-se um score. A avaliação é uma tarefa desgastante para o docente, em especial quando o número de alunos é grande, por exemplo, em ambientes virtuais de ensino para Ensino a Distância (EAD) em larga escala. Nestes ambientes o processo de avaliação é ainda mais complexo, pois o professor e alunos estão distantes um do outro. Ambientes virtuais tais como Teleduc [17] e Moodle [10] têm procurado oferecer serviços com correção automática de questões objetivas para amenizar os problemas da avaliação em EAD.

Por outro lado a modalidade de ensino a distância tem mostrado uma taxa de crescimento (cerca de 20% ao ano) bem superior à modalidade presencial, assim reforça-se a relevância de estudar avaliação automática em ambientes virtuais. Este trabalho explora o desafio de avaliar automaticamente Mapas Conceituais como questões abertas em ambientes virtuais de ensino.

Pesquisas realizadas com o intuito de criar ferramentas de avaliação automática de ensaios ocorrem desde meados de 1960, porém seus resultados eram incipientes, como exemplo pode-se citar o sistema PEG (Project Essay Grader) [5] que pertence àquela época. Em 1982, com a ferramenta *Writer's Workbench* foi dado o primeiro passo no sentido de solucionar as limitações da avaliação de ensaios [8]. Finalmente, em 1990, novas ferramentas e técnicas computacionais, baseadas em Processamento de Linguagem Natural (PLN) e Extração de Informação (EI), passaram a ser aplicadas buscando melhorar a performance da avaliação automática de questões discursivas [8]. Vale citar duas ferramentas a IEA (*Intelligent Essay Assess*) e o E-Rater [8]: O IEA utiliza LSA (*Latent Semantic Analysis*), uma técnica que permite comparações de

similaridade semântica de partes de um texto, alcançando uma acurácia de em média 85% em relação a avaliadores humanos [6]; O E-Rater utiliza PLN para identificar características lingüísticas tais como: sinais sintáticos e léxicos específicos. Sua abordagem alcança uma acurácia de 97% em relação a avaliadores humanos [4].

Com relação à avaliação manual, a avaliação automática oferece algumas vantagens tais como:

- Avaliação com retorno imediato do resultado para o aluno; fato relevante principalmente num ambiente de ensino a distancia onde o aluno não tem oportunidade de um contato com o professor;
- Liberação parcial da atividade de correção do professor; dando a oportunidade para o professor focalizar mais em atividades de orientação pessoal para alunos com problemas;
- Supervisão mais fina do aprendizado dos alunos (por exemplo, através da observação de relatórios comparativos entre os alunos de uma turma); permitindo ao professor identificar problemas mais cedo e tomar decisões para sua correção.

Além disso, outra vantagem pedagógica é que a avaliação automática de questões abertas contribui no ensino dos níveis mais altos da taxonomia de Bloom [3], que envolve trabalhos de análise e síntese. A taxonomia de Bloom (Figura 1) provê a divisão de questões em seis níveis de abstração: as questões fechadas (ou objetivas) contribuem com os níveis mais baixos da taxonomia, enquanto as questões abertas contribuem com níveis mais altos da taxonomia de Bloom [3].



Figura 1: Níveis da taxonomia de Bloom.

As questões abertas (não objetivas) são representadas por pequenos ensaios, questões discursivas e Mapas Conceituais. Neste trabalho propomos um método para avaliar Mapas Conceituais.

Este artigo está constituído de 7 seções: a seção 2 relata a fundamentação acerca dos Mapas Conceituais; a seção 3 descreve os trabalhos relacionados à avaliação discente, bem como avaliação de MCs; a seção 4 descreve o cenário da proposta; a seção 5 mostra os resultados obtidos; a seção 6 apresenta as conclusões; a seção 7 apresenta as referências bibliográficas.

2. Fundamentação Teórica: Mapas Conceituais

O Mapa Conceitual (MC) como abordagem de ensino foi proposta por Novak na década de 70 [13][12]. Com o MC, Novak buscava uma metodologia para aplicar na prática a Teoria da Aprendizagem Significativa, desenvolvida por Ausubel [2].

A Teoria da Aprendizagem Significativa tem como princípio a importância do aprendizado prévio de um aprendiz, ou seja, todo novo conhecimento adquirido deve ser relacionado a algum conceito presente na estrutura cognitiva do indivíduo, de modo que esse novo conhecimento seja atrelado a este conceito prévio. Essa ancoragem tem como resultado não somente a aquisição de novo conceito, mas também a ocorrência de mudanças no sentido de que tanto o novo conceito como o anterior a qual ele está sendo atrelado assumem novos significados.

A abordagem de MC busca externalizar o conhecimento presente na estrutura cognitiva do indivíduo através da representação dos conceitos e as relações entre eles. Assim como a estrutura cognitiva é construída de forma hierárquica, um Mapa Conceitual apresenta uma estrutura hierárquica que representa o conhecimento de um indivíduo através de conceitos e palavras de ligação, que são as relações entre os conceitos. A composição entre conceito-link-conceito é denominada proposição. O conjunto de proposições representa o conhecimento do indivíduo a respeito de determinado assunto [16][11]. A representação de Mapas Conceituais dá-se através de formas geométricas. Retângulos ou elipses representam os conceitos. Setas representam palavras de ligação entre dois conceitos, ver Figura 2.

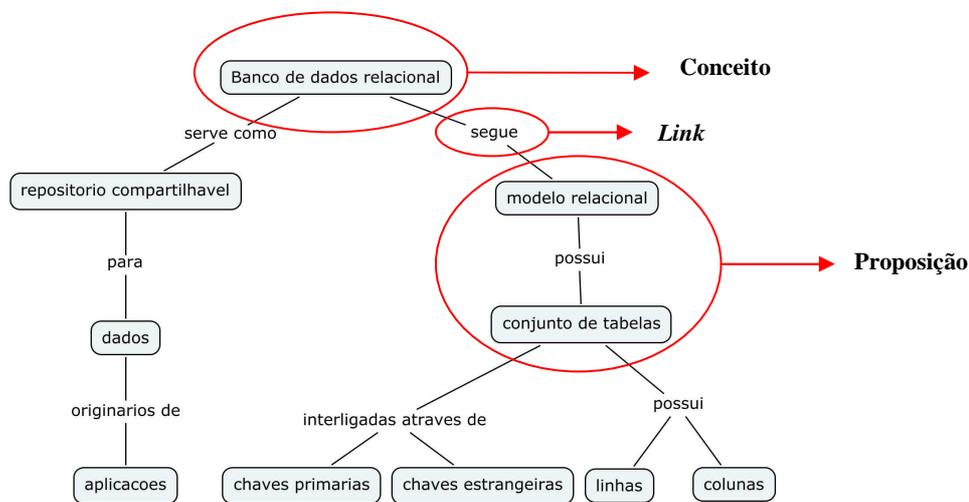


Figura 2: Exemplo com os conceitos sobre MC

3. Trabalhos Relacionados

Pesquisas relacionadas à avaliação automática de MCs são recentes. A proposta de [7] avalia MCs dos estudantes baseada na comparação contra o MC do especialista, de acordo com critérios tais como o número de conceitos, número de relacionamentos corretos, etc. Como resposta, é retornado um relatório qualitativo

informando conceitos errôneos e unidades de ensino que concernem ao assunto abordado no MC, auxiliando o aluno desenvolver uma nova resposta correta [7] .

A proposta de [1] busca avaliar a aprendizagem gradual do discente através do ensino orientado, onde os MCs desenvolvidos pelos alunos passam por etapas de correção em que o aluno refina seu MC sobre determinado assunto. Ao final do refinamento, o MC do aluno é submetido à avaliação, onde um agente realiza uma comparação com um modelo do professor. Este agente tenta adequar a resposta de acordo com padrões de resposta de alunos existentes em sua base de conhecimento; estes padrões catalogam formatos de resposta de alunos e à medida que a resposta do aluno se molda a um padrão, o agente atribui um determinado número pontos.

Também é proposta em [14][15] uma abordagem qualitativa para avaliar MCs, onde o MC do estudante é comparado contra uma Ontologia do assunto abordado. O retorno é um relatório detalhado mostrando se o conhecimento do aluno é ou não compartilhado com a ontologia. O forte da abordagem é a possibilidade de se fazer inferências sobre os conceitos do MC do estudante, para uma avaliação mais fina. Em contrapartida uma desvantagem é a necessidade de se compor uma detalhada ontologia para cada assunto sendo abordado.

Todos os três trabalhos [1][7] e [14] focam em proporcionar uma avaliação adaptativa, explorando mais detalhes de como desenvolver agentes que auxiliam ao estudante nas atividades de construção do conhecimento na edição dos MCs. Porém, nenhum deles detalha a abordagem de avaliação de um MC com o objetivo de produzir um escore, por exemplo, numa escala de 0 a 10. Em nosso trabalho focamos uma abordagem de avaliação quantitativa que produz um escore e validamos o processo de avaliação automática com relação a avaliadores humanos. Num trabalho futuro devemos explorar como melhorar o relatório qualitativo, de forma similar aos sistemas mencionados.

Cada um dos três sistemas exige a codificação de uma base de conhecimento (ou ontologia de domínio) com as informações necessárias para gerar a avaliação adaptativa. A codificação deste conhecimento não é um processo fácil. Em contrapartida, nossa abordagem trabalha apenas com um MC do professor: a “expansão do vocabulário” é feita a partir das melhores respostas dos estudantes.

4. Proposta

A proposta de avaliação automática para MCs tem como principal objetivo analisar a similaridade dos MCs dos estudantes contra um modelo de resposta do especialista, retornando um escore da avaliação quantitativa e um relatório da avaliação qualitativa. O objetivo é comparar, via n-gramas, o MC do estudante contra um modelo de resposta formada por vários MCs. A análise de bigramas leva em consideração a formação conceito-*link* ou *link*-conceito enquanto que a análise de trigramas leva em consideração a formação dos arcos conceito-*link*-conceito ou *link*-conceito-*link*. Antes de aplicar os n-gramas, os conceitos e palavras de ligação são pré-processados e normalizados via um *stemmer* e um dicionário de sinônimos.

O estudo de caso tem por base um conjunto de aproximadamente 400 MCs que são respostas de duas questões coletadas no ambiente de EAD chamado LabSQL [9], que é um ambiente virtual para ensino de programação SQL e Banco de Dados Relacional. Estes 400 MCs já avaliados por especialistas formam a base de treinamento e teste para o desenvolvimento dos modelos de avaliação.

O objetivo da avaliação é a partir da comparação produzir uma nota (nota) e um relatório diagnóstico. Na Figura 3 é mostrado um MC modelo de resposta e um mapa de um aluno sobre o conceito de Banco de dados relacional. O MC modelo pode ser lido como: “Um banco de dados relacional serve como repositório compartilhável para dados originários de aplicações; segue o modelo relacional formado por conjunto de tabelas; o conjunto de tabelas possui linhas e colunas. Um banco de dados relacional é composto por um conjunto de tabelas interligadas por chaves primárias e chaves estrangeiras”.

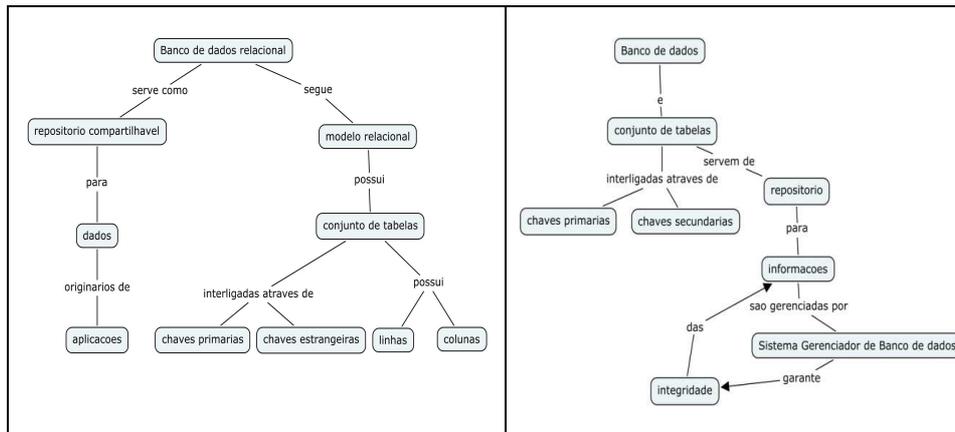


Figura 3: MC Modelo de resposta (à esquerda) e exemplo de MC de aluno (à direita).

O processo de avaliação segue quatro passos numa arquitetura de *pipeline*, onde o terceiro passo é utilizado para gerar um relatório de anomalias encontradas e o quarto passo é utilizado para calcular a nota:

- Primeiro: compara-se o MC de cada aluno com um MC modelo para identificar os melhores MCs;
- Segundo: cria-se modelo de resposta com **vocabulário expandido**, composto pelo mapa modelo do professor mais os 2 a 4 MCs escolhidos pelo primeiro passo;
- Terceiro: gera-se um relatório de diagnóstico e orientação comparando-se o MC do aluno contra o modelo com vocabulário expandido: identifica-se todos os bigramas e trigramas que estão num modelo e não estão no outro;
- Quarto: gera-se uma nota utilizando o método de classificação KNN (K vizinhos mais próximos); compara-se cada MC de um aluno com o modelo de vocabulário expandido registrando-se as métricas de bigramas e trigramas; a partir das medidas escolhe-se os K-vizinhos mais próximos para produzir a nota.

No terceiro passo identificam-se todos os trigramas (conceito-*link*-conceito ou *link*-conceito-*link*) que estão no MC do estudante e que não estão no MC modelo de resposta e vice-versa. Idem para os Bigramas. Esta informação deve ser utilizada para fazer um relatório de diagnóstico para o estudante, o problema é como utilizar esta

informação (o que dizer) para gerar um relatório que faça com que o estudante aprimore o seu MC. Se for mostrado tudo o aluno não fará mais esforço para aprimorar a sua solução. Uma solução pode ser incluir “ruído” junto nas recomendações ao aluno para ele ter que fazer suas próprias escolhas pelo certo.

Como exemplo de informação do relatório de diagnóstico, com base nos MC da Figura 3, listamos os trigramas conceito-*link*-conceito não existentes no MC do estudante: [modelo relacional possui conjunto de tabelas][Banco de dados relacional serve como repositório compartilhável][dados originários de aplicações] [conjunto de tabelas possui linhas] [conjunto de tabelas possui colunas].

5. Resultados Obtidos

Os MCs gerados pelos alunos foram submetidos à análise de n-gramas. Foram realizados vários experimentos para identificar quais métricas de n-gramas deveriam ser consideradas: as melhores métricas foram bi-grama e tri-grama.

Os experimentos tiveram como base modelos, cujas respostas utilizavam expansão de vocabulário, ou seja, o modelo de resposta era formado por mais de um MC. A Figura 4 mostra a média de erros variando-se o número de MC, de 1 até 4, para compor a resposta modelo. A partir de 4 MCs a taxa de erros cresce. Por outro lado, os experimentos para o KNN mostraram que o melhor valor de K é 3.

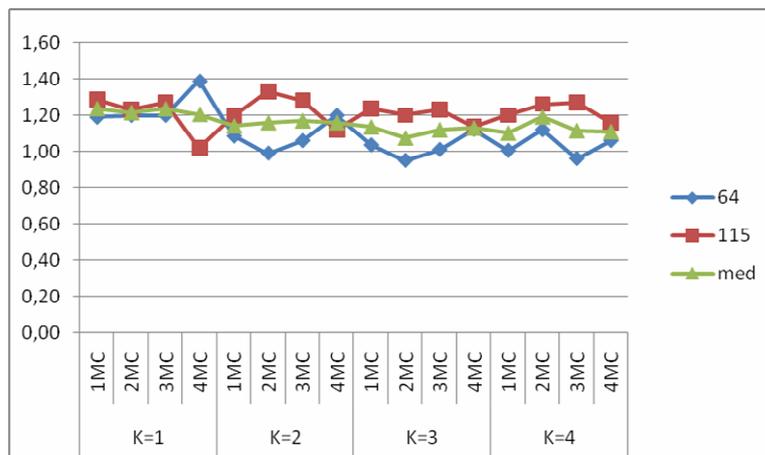


Figura 4: Média de erros para as questões 64 e 115, variando-se o número de MC (1 até 4) para compor a resposta modelo e o valor do K (1 até 4) para KNN: menor erro 1,08 com 2 MC e K=3.

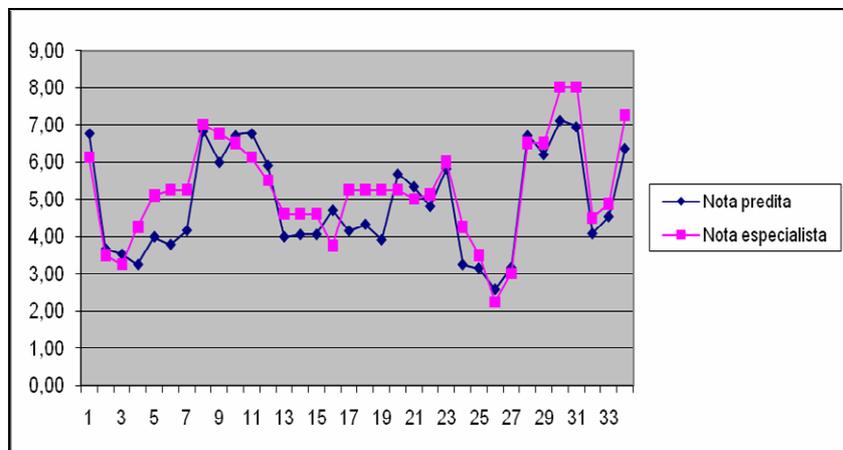


Figura 5: Notas do especialista comparadas com as notas preditas pelo sistema para a questão 64, para reposta modelo com 2 MC e para o classificador KNN com K=3.

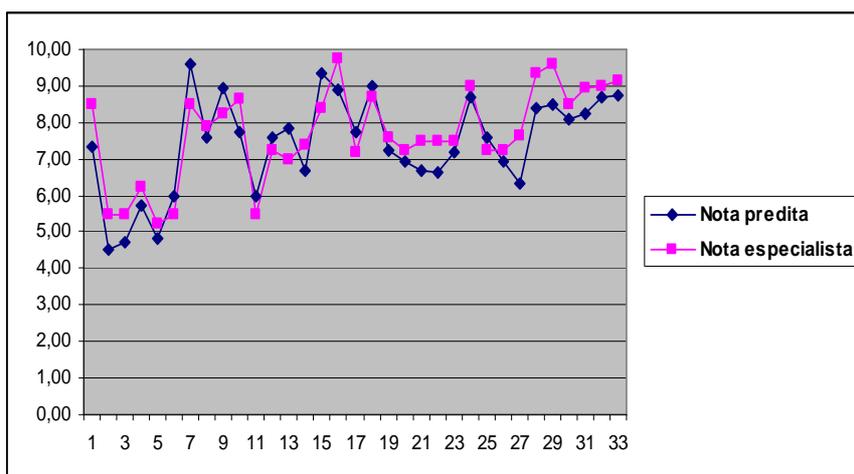


Figura 6: Notas do especialista comparadas com as notas preditas pelo sistema para a questão 115, para reposta modelo com 2 MC e para o classificador KNN com K=3.

As Figuras 5 e 6 mostram a diferença entre as notas preditas pelo sistema e as notas atribuídas por um especialista humano para as questões 64 e 115, respectivamente, agora utilizando-se apenas os melhores parâmetros para expansão do vocabulário e para o classificador KNN (2 MC e K=3). As médias de acertos foram de 90% e 89% respectivamente.

6. Conclusão

A avaliação discente não é uma tarefa trivial sem o auxílio de software para a sua automação. A abordagem de MCs é um método de avaliação de questões abertas. Pode-se avaliar um aluno tanto quantitativamente, através de uma nota, como qualitativamente, evidenciando as dificuldades do discente.

A tarefa de predição de dados é amplamente estudada e assim como a técnica utilizada neste artigo há outras abordagens que seriam possíveis na realização destes experimentos, tais como a técnica estatística de Regressão Linear, a utilização de Redes Neurais, entre outros. Entretanto, o KNN foi escolhido por sua simplicidade e por sua fácil adaptação a dados de valores contínuos.

Este experimento teve como finalidade verificar a efetividade de técnicas de IA como técnicas de avaliação quantitativa de MCs. Os resultados obtidos mostram que o problema de apenas avaliar em termos de nota é fácil de ser resolvido, obtivemos uma média de acertos de 90% em relação às avaliações realizadas por avaliadores humanos. Em nossos experimentos a utilização de dicionário de sinônimos não provê uma melhora significativa na acurácia ao predizer a nota para um MC, entretanto é útil para mostrar um relatório das triplas Conceito-*Link*-Conceito ausentes nos MCs dos alunos.

É necessário ponderar que a avaliação proposta compreende duas etapas uma quantitativa e outra qualitativa. Percebemos ser importante um relatório mostrando o que não deve estar no MC do estudante. Por outro lado também, percebemos que não se pode disponibilizar um relatório detalhado, exibindo tudo o que foi omitido no MC do aluno, pois assim estaríamos respondendo a questão por ele. Estamos considerando também a inclusão de “ruído” no relatório de diagnóstico forçando o aluno a fazer escolhas. Em resumo, estamos trabalhando numa proposta de um relatório que visa auxiliar (orientar) ao aluno a melhorar a sua resposta permitindo múltiplas submissões imediatas, porém procurando manter o necessário esforço individual no processo de aprendizagem.

Como trabalho futuro, nesta linha de avaliação automática com múltiplas submissões de uma resposta, está o desenvolvimento de uma metodologia para trabalhar com avaliação adaptativa centrada em permitir um maior alinhamento entre o nível de conhecimento atual do aprendiz com as tarefas que ele deve executar para adquirir gradualmente o aprendizado necessário. Esse refinamento deve ser alcançado através da inclusão de “ruídos” no relatório de diagnóstico. Para incluir estes “ruídos” é necessário gerá-los automaticamente, combinando informações de MCs com os menores escores juntamente com as informações provenientes do MC modelo de resposta.

Referências

1. Anohina A., Graudina V., Grundspenkis J. Using Concept Maps in Adaptive Knowledge Assessment. Proceedings of the 15th International Conference on Information Systems Development “Methods and Tools, Theory and Practice”, August 31 - September 2, 2006, Budapest, Hungary.
2. Ausubel, D. Psicologia Educacional. - 2. Ed. – Rio de Janeiro: Interamericana. 1980.
3. Bloom B. S. Engelhart M, Furst E, Hill W, And Krathwohl D. R., 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of

college and university examiners. Handbook I: Cognitive Domain. New York: Longmans, Green.

4. Burstein, J. (2003). The e-rater scoring engine: Automated Essay Scoring with natural language processing. In M. D. Shermis and J. C. BURSTEIN (Eds.), *Automated Essay Scoring: A cross disciplinary approach* (pp. 113–121). Mahwah, NJ: Lawrence Erlbaum Associates.
5. Dikli, Semire. 2006. An Overview of Automated Scoring of Essays, *JTLA - The Journal of Technology, Learning, and Assessment*, vol. 05, no. 1, August 2006.
6. Foltz, P. W. (1996). Latent Semantic Analysis for text-based research. *Behavior Research Methods, Instruments and Computers*, 28(2), 197–202. Retrieved June 6, 2004 from <http://www-psych.nmsu.edu/~pfoltz/>
7. Gouli, Evangelia; Gogoulou, Agoritsa; Papanikolaou, Kyparisia; Grigoriadou, Maria. *Compass: An Adaptive Web-Based Concept Map Assessment Tool*. Proc. of the First Int. Conference on Concept Mapping. Pamplona, Spain. 2004
8. Hearst M.A. The Debate on Automated Essay Grading. *IEEE Intelligent Systems*, vol. 15, no. 5, pp. 22-37, September/October. 2000.
9. Lino, A. D. P.; Silva, A. S.; Santos, T.L.T.; Harb, M.P.A.H.; Favero, E.L.; Brito, S.R. *Avaliação automática de consultas SQL em ambiente virtual de ensino-aprendizagem*. Conferencia Ibérica de Sistemas y Tecnologías de la Información. CISTI 2007.
10. Moodle (2008). Disponível Em: <[Http://Moodle.Org/](http://Moodle.Org/)>. Acesso Em: 18 Abr. 2008
11. Moreira, M.; Buchweitz, B. *Mapas Conceituais – Instrumentos Didáticos e Análise de Currículo*. São Paulo: Moraes. 1987.
12. Novak, J.D. Cañas, A.J. (2006). The theory underlying concept maps and how to construct them. Technical Report IHMC Cmap Tools 2006-1, Florida Institute for Human and Machine Cognition. Available online at: <http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf>
13. Novak, J. D. Gowin, D. B. (1984). *Learning how to learn*. New York and Cambridge, UK: Cambridge University Press.
14. Rocha, F. E. L. *Avaliação da Aprendizagem: Uma Abordagem Qualitativa Baseada em Mapas Conceituais, Ontologias e Algoritmos Genéticos*. PhD Thesis. Centro Tecnológico, Federal University of Pará, Brazil. 2007.
15. Rocha, F. E. L., Costa Jr, J. V. & Favero, E. L. *Qualitative Learning Assessment via Concept Maps*. In: *Revista de Informática Teórica e Aplicada*, 14(1). 2007.
16. Ruiz-Primo, M. A. Shultz, S. E., Li, M., Shalveson, R. J. (2001). Comparison of the reliability and validity of notas from two concept-mapping techniques. *Journal of Research in Science Teaching*, 38(2), 260-278.
17. Teleduc (2008). Disponível em: <<http://teleduc.nied.unicamp.br/teleduc>>. Acesso em: 23 mai. 2008