

Um QAsystem para Interação de Alunos em Avaliações Somativas a Distância

Márcia G. de Oliveira¹, Elias Oliveira², Rafael Z. Marchesi³

¹ Programa de Pós Graduação em Engenharia Elétrica (PPGEE)
Universidade Federal do Espírito Santo (UFES)
Caixa Postal 5011 – 29.060-970 – Vitória – ES – Brazil

² Departamento de Ciências da Informação – UFES
Vitória – ES – Brazil

³ Programa de Pós Graduação em Informática (PPGI) – UFES
Vitória – ES – Brazil

clickmarcia@gmail.com, elias@acm.org, rzanolim@yahoo.com.br

Abstract. *This paper presents a QAsystem, an online system of questions and answers, which will return an appropriate response to a student if the system knowledge base contains compatible questions to question asked. For this, the QAsystem, which is based on clustering, analyzes that your questions are similar to the basic question of the student returning the answers relating to questions of greatest similarity. Tests conducted to evaluate the effectiveness of QAsystem to recognize similar questions of an international base of questions, the TREC-9, point to results more than 85%, a considerable achievement compared to other experiences with this type of system.*

Keywords: *QAsystem, Clustering, Summative assessment, Classification of documents*

Resumo. *Este trabalho apresenta um QAsystem, isto é, um sistema online de perguntas e respostas, que retornará uma resposta adequada para um aluno caso exista na base de conhecimento desse sistema uma resposta compatível à pergunta realizada. Para isso, o QAsystem, que é baseado em clusterização automática, verificará quais perguntas de sua base são similares à pergunta do aluno, retornando as respostas referentes às perguntas de maior similaridade. Os testes realizados para avaliação da eficácia do QAsystem em reconhecer perguntas similares de uma base internacional de perguntas, a TREC-9, apontam para resultados superiores a 85%, um resultado considerável comparado a outras experiências com esse tipo de sistema.*

Palavras-chave: *QAsystem, Clusterização automática, Avaliação somativa, Classificação de documentos*

1. Introdução

A avaliação somativa deve ser um meio de levantar balanços confiáveis ao final de um processo de ensino-aprendizagem [Ballester 2003]. Em cursos a distância, a avaliação

somativa, no entanto, ainda é um forma de avaliação que gera muitas desconfianças e resistências, embora seja muito valorizada no Ensino Presencial [Maia 2001].

Uma forma de reduzir as resistências à avaliação somativa em cursos de Educação a distância é deslocar os seus propósitos de classificar alunos e de supervalorizar o erro para a qualificação de um processo de ensino-aprendizagem. Isso significa dar à avaliação somativa *online* o papel de consolidar um processo avaliativo bem-sucedido iniciado por avaliações diagnóstica e formativa [Oliveira 2009].

O sistema SOAD (*Sistema Online de Avaliação a Distância*) é um sistema de avaliação somativa projetado como extensão de um sistema de avaliações diagnóstica e formativa, o SOAC¹ (*Sistema Online de Atividades de Classificação*) [Oliveira 2009]. A ideia de avaliação somativa do sistema SOAD é aferir a qualidade de um processo formativo de aprendizagem. Assim como o SOAC, o sistema SOAD está sendo desenvolvido com a finalidade de apoiar o processo avaliativo do ensino-aprendizagem de classificação bibliográfica em cursos de Biblioteconomia.

Uma importante característica do modelo de avaliação somativa do SOAD é possibilitar que um aluno faça perguntas durante uma prova *online*. Isso incentiva o aluno a desenvolver a habilidade de questionar e o ajuda a chegar à resposta certa caso saiba fazer as perguntas certas. De acordo com [Antunes 2001], fazer perguntas relevantes do tipo *o que, quem, onde, quando, como e por que* expressam passos que nem todos sabem dar.

Com o objetivo de promover maior interatividade em provas *online*, o sistema SOAD disponibilizará um *QAsystem*, isto é, um sistema que responderá automaticamente as perguntas feitas por alunos, caso tenha respostas para essas perguntas. Fazendo isso, o processo de avaliação *online* se torna mais dinâmico e eficaz, devido ao possível *feedback* imediato dado ao aluno.

Para responder uma pergunta de um aluno, o *QAsystem* verifica se há em sua base de conhecimento perguntas similares à pergunta feita por esse aluno. Se sim, o aluno receberá automaticamente como resposta à sua pergunta as respostas das perguntas similares reconhecidas pelo *QAsystem*. Caso contrário, a pergunta do aluno será arquivada para análise do professor. Se o professor considerar a pergunta relevante e a responder, a pergunta será adicionada à base de perguntas do *QAsystem*.

O *QAsystem* do SOAD é baseado em clusterização automática, que é uma técnica de reconhecimento de padrões que consiste em agrupar padrões com base em suas características similares [Fahim et al. 2006]. Para medir a eficácia do *QAsystem* em reconhecer perguntas similares por clusterização automática, foram realizados testes em uma base de perguntas internacional, a TREC-9. Os resultados obtidos nas principais métricas de avaliação de clusterização foram superiores a 85%, o que valida a clusterização como uma técnica eficaz para o funcionamento de um *QAsystem* no sistema SOAD.

Este trabalho está organizado na ordem descrita a seguir. Na Seção 2, é apresentado o sistema de avaliação somativa SOAD. Na Seção 3, é descrita uma metodologia de *QAsystem* livre de gramática baseada em clusterização automática. Na Seção 4, são discutidos os testes e os resultados de avaliação do *QAsystem* proposto. Na Seção 5, são apresentadas as considerações finais e os trabalhos futuros.

¹Disponível em : <http://www.informarcia.pro.br/soac>

2. O Sistema SOAD

O *SOAD* é um sistema de avaliação somativa *online* do processo de ensino-aprendizagem de classificação bibliográfica em cursos de Biblioteconomia.

O *SOAD* tem como objetivo aferir, através de provas *online*, a qualidade de um processo de ensino-aprendizagem e possivelmente apontar competências de alunos. A avaliação somativa do *SOAD* foi projetada para ser realizada após as avaliações diagnósticas e formativas do sistema *SOAC* [Oliveira 2009].

O sistema *SOAD* ainda está em fase de desenvolvimento e está inicialmente hospedado no seguinte endereço: <http://www.informarcia.pro.br/soad>.

Quando um aluno inicia uma avaliação *online*, as questões de sua prova são sorteadas. O aluno resolve cada questão de uma prova através da tela da Figura 1.

SOAD

Sistema Online de Avaliação a Distância

:: Prova : <CÓDIGO> :: Aluno: <Nome Aluno> :: Nível: <número do nível> :: Tempo : 02:00:00

Questão <Nº> - Que código(s) de classificação bibliográfica você atribuiria para a descrição a seguir?

[Exemplo de Objeto](#)

Códigos

Adicionar >>

Não sei responder

Apagar

Apagar

:: Deseja fazer perguntas?

[Voltar](#) [Avançar](#)

Salvar Finalizar

[?](#) [Ajuda](#)

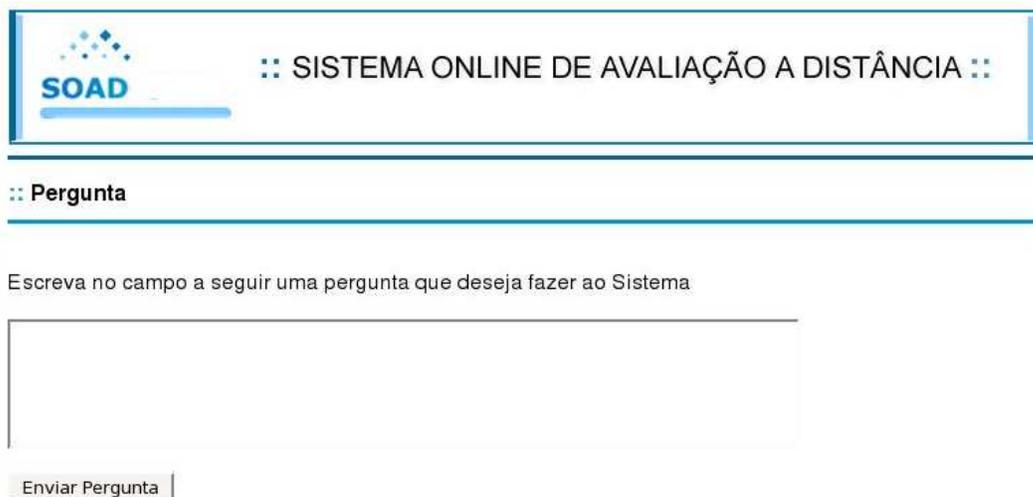
Figura 1. Resolução das questões de uma prova *online*

De acordo com a Figura 1, o aluno visualiza uma série de informações como o código da prova, o nível de dificuldade da questão e a contagem do tempo, que são definidas pelo professor na especificação das provas.

Para cada questão de uma prova *online*, o processo de resolução é como o do sistema *SOAC* [Oliveira 2009], isto é, consiste em atribuir um ou mais códigos de classificação bibliográfica a um documento através do botão **Adicionar** (Ver Figura 1). Esse documento é apresentado na forma de um *link* para um documento que pode ser um texto, uma imagem, um artigo ou qualquer item de informação que possa ser classificado.

Após essa visão geral do sistema SOAD e de suas principais funcionalidades fornecidas para o aluno, pode-se explicar a funcionalidade que representa o foco deste trabalho que é o sistema de perguntas e respostas.

Na resolução de uma questão, de acordo com a Figura 1, o aluno também poderá fazer perguntas sobre um objeto de classificação clicando em **Perguntar**. Quando um aluno faz uma pergunta (ver Figura 2) sobre um objeto de classificação, essa pergunta será comparada com outras perguntas associadas a esse mesmo objeto por um *QAsystem*. Caso reconheça perguntas similares, o *QAsystem* apresentará para o aluno as respostas das perguntas similares. Caso contrário, a pergunta será arquivada no sistema para ser analisada, e talvez respondida, pelo professor. Se for respondida, a pergunta será inserida na base de perguntas que forma o conhecimento do *QAsystem*.



A imagem mostra a interface do sistema SOAD. No topo, há um cabeçalho com o logo "SOAD" à esquerda e o texto ":: SISTEMA ONLINE DE AVALIAÇÃO A DISTÂNCIA ::" à direita. Abaixo do cabeçalho, há uma seção intitulada ":: Pergunta". O texto principal da seção diz: "Escreva no campo a seguir uma pergunta que deseja fazer ao Sistema". Abaixo desse texto, há um campo de entrada de texto vazio. Na base do campo, há um botão com o texto "Enviar Pergunta".

Figura 2. Fazer perguntas

Para cada questão de prova que um aluno resolve, são considerados na avaliação de desempenho a atribuição de códigos de classificação bibliográfica e as perguntas feitas por esse aluno. Caso o aluno tenha optado por não atribuir códigos, a sua justificativa, ao ser analisada pelo professor, poderá também ser pontuada.

3. Uma proposta de *QAsystem* para o sistema SOAD

Um *QAsystem* tem o papel de interagir com usuário fornecendo respostas às suas perguntas. Muitas propostas de *QAsystems* seguem essencialmente as abordagens livres de gramática [Cui et al. 2005], híbridas [Achananuparp et al. 2008] ou baseadas em processamento de linguagem natural como os *chatterbots* [Teixeira et al. 2005].

Uma abordagem híbrida de *QAsystem* é proposta por [Achananuparp et al. 2008]. Essa abordagem combina o processamento de linguagem natural com técnicas de recuperação da informação. Dessa forma, compõe-se um medida híbrida de similaridades através de componentes semânticos, sintáticos e de classificação de tipos de questões.

Neste trabalho, ao contrário das abordagens de [Achananuparp et al. 2008] e [Teixeira et al. 2005], é proposta uma abordagem de *QAsystem* livre de gramática, que

aplica a clusterização automática para reconhecer perguntas semelhantes através de medidas tradicionais de similaridade de textos.

3.1. A Clusterização automática

A clusterização é um técnica de reconhecimento de padrões que utiliza a aprendizagem não-supervisionada para realizar o agrupamento, isto é, a auto-organização de padrões em classes não-rotuladas (*clusters*) [Jain et al. 1999].

O *K-means* é uma das técnicas mais aplicadas na área de classificação de padrões. Para este trabalho, foi escolhida a técnica *Bisecting K-means*, uma variação do *K-means* que funciona da seguinte forma [Steinbach et al. 2000]:

1. Dividir um *cluster* em dois *sub-clusters* utilizando o *K-means* [Looks et al. 2007]
2. Repetir o passo 1 por *I* vezes e escolher a melhor divisão por similaridade.
3. Repetir os passos 1 e 2 até alcançar o número de *clusters* desejado.

Os padrões se auto-organizam em *clusters* com base em medidas de similaridades. Para os experimentos deste trabalho foram escolhidas as medidas *coseno* e *correlação*. A medida de similaridade *coseno* é calculada pelo produto interno entre dois vetores *A* e *B*, que representam as características de dois padrões, da seguinte forma:

$$Sim(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|}, \text{ onde } \theta \text{ é o ângulo entre } A \text{ e } B.$$

O *coeficiente de correlação* mede o grau de relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. O valor 0 (zero) significa que não há relação linear, o valor 1 indica uma relação linear perfeita e o valor -1 também indica uma relação linear perfeita mas inversa, ou seja quando uma das variáveis aumenta a outra diminui. Dadas duas variáveis ou vetores *x* e *y*, calcula-se o coeficiente de correlação *r* da seguinte forma:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \text{ onde } \bar{x} \text{ e } \bar{y} \text{ são vetores-médias de } x, y$$

Existem vários critérios para avaliar a clusterização. Neste trabalho, foram escolhidas as tradicionais métricas *recall*, *precision* e *F₁* e a métrica *Rand Index* [Jain et al. 1999, Manning et al. 2008].

Para a realização da clusterização automática foram utilizados os algoritmos do *software Cluto 2.1.2*². O *Cluto* é um pacote de *software* para clusterização de conjuntos de dados de baixas e altas dimensões. A característica principal desse *software* é que seus algoritmos tratam o problema da clusterização como um processo de otimização que busca maximizar ou minimizar uma particular função-objetivo de clusterização.

3.2. Funcionamento do QAsystem

A Figura 3 é uma representação do modelo funcional do *QAsystem* proposto neste trabalho para responder automaticamente às perguntas feitas por alunos em avaliações somativas do sistema SOAD.

De acordo com a Figura 3, um aluno faz uma pergunta no sistema SOAD que é recebida pelo *QAsystem*. Em seguida, essa pergunta é submetida a um processo de indexação, que consiste em compor uma representação *P* mais compacta da pergunta através

²Software e documentação disponíveis em : <http://glaros.dtc.umn.edu/gkhome/views/cluto>

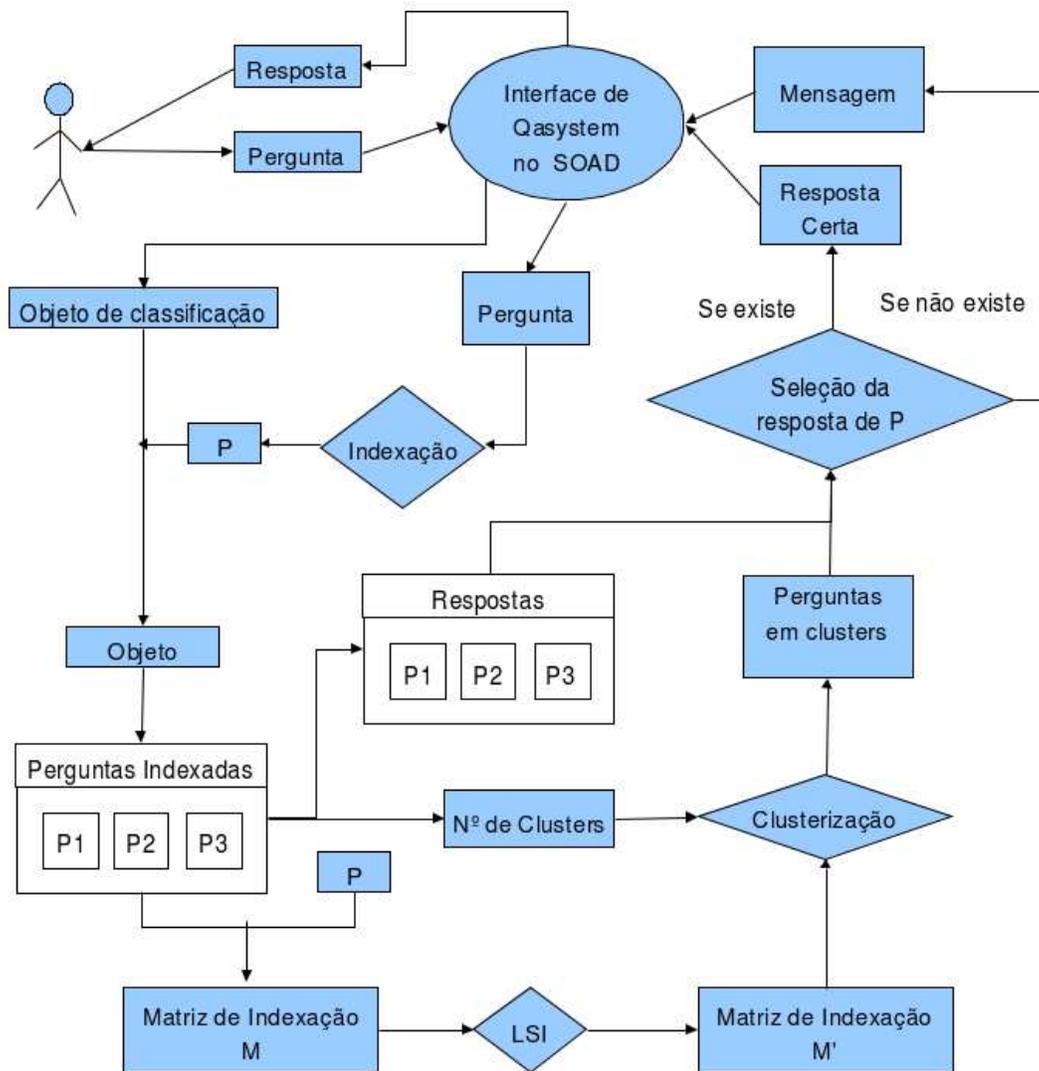


Figura 3. Funcionamento do *QAsystem* do SOAD

de um conjunto de termos ou palavras que compõem a pergunta original. A pergunta P e as perguntas associadas a um objeto de classificação, já indexadas também, são reunidas em uma Matriz de Indexação M . A Matriz M e o número de *clusters* que se deseja formar são fornecidos como entradas dos algoritmos de clusterização do *software Cluto*. Após o processo de clusterização, as perguntas são agrupadas em k *clusters* de acordo com os índices de similaridade (cosseno ou correlação) mais próximos. As respostas das perguntas, que estão no mesmo *cluster* da pergunta P , são enviadas como respostas para o aluno. Caso não haja perguntas similares no mesmo *cluster* da pergunta P , é enviada para o aluno uma mensagem de que não foi encontrada uma resposta para a sua pergunta.

4. Experimentos e Resultados

Para [Achananuparp et al. 2008], apenas as tradicionais técnicas de similaridades de documentos não são eficazes quando o tamanho de uma sentença é muito curto e possui poucas palavras em comum com outras sentenças. Este trabalho, no entanto, mostra

que, apenas com as tradicionais medidas de similaridade de documentos como *coseno* [Baeza-Yates and Ribeiro-Neto 1998] e correlação, que são livres de gramática, a clusterização automática produz resultados muito bons para um *QAsystem*.

4.1. Metodologia

A metodologia de desenvolvimento do *QAsystem* para o sistema *SOAD* consistiu das seguintes etapas:

Escolha do conjunto de dados

Para esse experimento, foi utilizado o conjunto de perguntas da base *TREC-9*³. Esta base contém 247 perguntas e é de língua inglesa. As perguntas dessa base foram escolhidas porque foram construídas por assessores humanos para serem semanticamente idênticas, mas sintaticamente diferentes [Achananuparp et al. 2008], isto é, paráfrases.

Pré-processamento dos dados

Para serem submetidos ao processo de clusterização automática, a base de perguntas *TREC-9* passou pelas seguintes etapas de pré-processamento: separação em arquivos, retirada de *stopwords* e indexação.

A etapa de separação em arquivos foi realizada pela colocação de cada pergunta da base *TREC-9* em arquivos do tipo texto.

A etapa de retirada *stopwords* consistiu em retirar palavras que eram irrelevantes para a representação de informações textuais. Nessa etapa, foram retiradas dos arquivos de perguntas *TREC-9* as palavras que continham apenas duas letras ou que tivessem até cinco letras e começassem com *th*, como, por exemplo, *then*, *that* e *this*. Não retiramos, porém, palavras que caracterizassem perguntas como as palavras *who*, *what* e *when*.

Na etapa de indexação, os arquivos com as perguntas foram representados vetorialmente através de uma Matriz *M*. Nessa matriz, as linhas indicavam cada documento, isto é, cada pergunta da base *TREC-9*. Cada coluna, por sua vez, representava a frequência de um termo ou palavra em cada documento dessa base.

De acordo com a Figura 4, na primeira linha de um arquivo de entrada dos algoritmos de clusterização do *software Cluto*, foram colocados o número de documentos e o número de palavras da base de perguntas e, em seguida, as linhas da matriz *M*.

Execução de Testes

Para a realização dos testes, a matriz *M* construída na etapa de indexação foi submetida aos algoritmos de clusterização automática do *Software Cluto*. Os resultados desses testes são apresentados a seguir.

4.2. Resultados

A Tabela 1 apresenta os resultados de clusterização automática sobre as perguntas da base *TREC-9* para formar agrupamentos de perguntas similares.

De acordo com a Tabela 1, as métricas de *Precision*, *Recall*, *F₁* e *Rand Index* revelam bons resultados de clusterização automática para as medidas de similaridade de *coseno* (*cos*) e *coeficiente de correlação* (*corr*).

³Disponível em : http://trec.nist.gov/data/qa/T9_QAdata/variants.key

Combination of Measures	Recall	Combined Changes	Rejection	Combined Changes	F ₁	Combined Changes
<i>sv+wo</i>	0.95	-	0.76	-	0.84	-
<i>sv+ps</i>	0.95		0.98		0.96	
<i>ps+wo</i>	0.96		0.98		0.97	
<i>sv+wo+rank</i>	0.98	8.4%	0.93	25.4%	0.95	17.3%
<i>sv+ps+rank</i>	0.98		1.00		0.99	
<i>ps+wo+rank</i>	0.98		0.99		0.98	
<i>sv+wo+conf</i>	1.00	14.7%	0.69	-13.3%	0.82	0%
<i>sv+ps+conf</i>	1.00		0.98		0.99	
<i>ps+wo+conf</i>	1.00		0.94		0.97	
<i>Jaccard</i>	0.41	-	1	-	0.58	-
<i>TF-IDF</i>	0.67	-	1	-	0.80	-

Figura 5. Resultados de medidas de similaridades aplicadas à base TREC-9. Fonte: [Achananuparp et al. 2008]

[Achananuparp et al. 2008], conforme a Figura 5. No entanto, os resultados de clusterização automática não superam a maioria das abordagens híbridas, isto é, de combinação de medidas de similaridades, propostas por [Achananuparp et al. 2008].

Ainda assim, concluímos que a clusterização automática, com medidas de similaridades de sentenças livres de gramática, pode ser um bom método para construção de *QAsystems* e menos complexo que as abordagens híbridas de [Achananuparp et al. 2008].

5. Considerações finais

Apresentamos neste trabalho, como parte integrante do sistema de avaliação somativa SOAD, um de seus principais recursos que é um *QAsystem* para responder automaticamente às perguntas feitas por alunos sobre objetos de classificação em questões de provas *online*. O objetivo de um *QAsystem* no SOAD é ajudar alunos a chegarem às respostas certas desde que saibam fazer as perguntas certas.

Para validar o *QAsystem* baseado em clusterização automática, realizamos alguns experimentos sob uma base de perguntas e respostas internacionalmente reconhecida para testes de *QAsystems* [Achananuparp et al. 2008]. Esses experimentos foram realizados aplicando-se a clusterização automática para formação de agrupamentos de perguntas similares. Os resultados alcançados foram superiores a 85% nas principais métricas de avaliação de clusterização automática.

Concluímos, dessa forma, que a clusterização pode ser um método eficaz para a formação de um *QAsystem* no sistema SOAD. O próximo passo de nossas pesquisas será realizar esses experimentos sobre uma base de perguntas de língua portuguesa.

Em resumo, este trabalho se propõe a oferecer como contribuição para cursos a distância *online* uma possibilidade de maior interatividade em provas *online* de forma a ajudar o aluno a construir caminhos para chegar às soluções de questões de provas.

6. Agradecimentos

Agradecemos à Fundação de Apoio à Ciência e Tecnologia do Espírito Santo (FAPES) pelo apoio dado às pesquisas apresentadas neste artigo.

Referências

- Achananuparp, P., Hu, X., Zhou, X., and Zhang, X. (2008). Utilizing sentence similarity and question type similarity to response to similar questions in knowledge-sharing community. In *17th international conference on World Wide Web*, Beijing, China. International workshop on question answering on the web (QAWeb2008).
- Antunes, C. (2001). *Trabalhando habilidades: construindo idéias*. Scipione, São Paulo, SP.
- Baeza-Yates, R. and Ribeiro-Neto, B. (1998). *Modern Information Retrieval*. Addison-Wesley, New York, 1 edition.
- Ballester, M. (2003). *Avaliação como apoio à aprendizagem*. Artmed, Porto Alegre, RS. Trad. Valério Campos.
- Cui, H., Sun, R., Li, K., Kan, M.-Y., and Chua, T.-S. (2005). Question answering passage retrieval using dependency relations. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 400–407, New York, NY, USA. ACM.
- Fahim, A., Salem, A., Torkey, F., and Ramadan, M. (2006). An efficient enhanced K-means clustering algorithm. *Journal of Zhejiang University SCIENCE A*, 7(10):1626–1633.
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323.
- Looks, M., Levine, A., Covington, G., Loui, R., Lockwood, J., and Cho, Y. (2007). Streaming Hierarchical Clustering for Concept Mining. *Aerospace Conference, 2007 IEEE*, pages 1–12.
- Maia, R. (2001). *Safes: um servidor de avaliações formativas e somativas para o ensino online via web*. PhD thesis, Scipione, São Paulo, SP.
- Manning, C., Raghavan, P., and Schütze, H. (2008). *Introduction to information retrieval*. Editora Thomson, Cambridge University Press.
- Oliveira, M. (2009). *Avaliações online para nivelamento e formação de classificadores humanos*. Master's thesis, Universidade Federal do Espírito Santo (UFES), Vitória, ES.
- Steinbach, M., Karypis, G., and Kumar, V. (2000). A comparison of document clustering techniques. *KDD workshop on text mining*.
- Teixeira, S., Ramiro, B., Oliveira, E., and Menezes, C. (2005). Chatterbots em ambientes de aprendizagem: uma proposta para a construção de bases de conhecimento. In *Anais do XXV Congresso da Sociedade Brasileira de Computação*, São Leopoldo, RS. SBC.