

Auto-Avaliador Colaborativo e Inteligente de Respostas

Lucas B. Galhardi¹, Jacques D. Brancher¹

¹Programa de Pós-graduação em Ciência da Computação –
Universidade Estadual de Londrina (UEL)
Caixa Postal 10.011 – CEP 86057-970 – Londrina – Paraná – Brasil

{lucasbgalhardi, jacques}@uel.br

Resumo. *Avaliações são muito utilizadas nos contextos de aprendizagem para verificar quanto conhecimento está sendo retido pelos alunos. Questões discursivas podem avaliar níveis diferentes do aprendizado dos alunos, quando comparadas com questões de múltipla escolha. Entretanto, devido a sua facilidade na correção, questões de múltipla escolha são geralmente mais utilizadas. Visando auxiliar nesse problema e apresentar ao professor uma ferramenta que lhe permita aplicar avaliações com questões discursivas sem receio do tempo de correção, surge o Auto-Avaliador Colaborativo e Inteligente de Respostas, uma ferramenta para a avaliação automática deste tipo de questões.*

1. Cenário de Uso

Avaliações são recorrentemente utilizadas no ambiente de aprendizado para verificar o conhecimento retido pelos alunos. Apesar de sua importância, professores geralmente passam por dificuldades ao avaliar respostas discursivas de salas de aula lotadas. Muitas vezes o trabalho da correção tem que ser levado para casa, comprometendo a qualidade de vida do professor [Jacomini and Penna 2016]. Pesquisas indicam que cerca de 75% dos professores afirmam levar trabalho para casa com frequência, como a correção de atividades [Nascimento and Santos 2015]. Essa situação sobrecarrega o professor e, consequentemente, diminui seu tempo para a preparação de aulas e outras atividades, prejudicando todo o ambiente de aula [Silva and Rosso 2008].

A sobrecarga de trabalho também pode levar o professor a optar por mais questões de múltipla escolha ao invés de questões que exijam uma resposta discursiva. Entretanto, questões discursivas avaliam níveis de aprendizado muitas vezes não contemplados por questões de múltipla escolha [Burrows et al. 2015]. Como são muitos alunos para apenas um professor, se questões discursivas forem utilizadas, ocasionará uma demora para os alunos obterem o *feedback* apropriado. Além disso, quando os alunos finalmente obtiverem sua avaliação, ela pode ser diferente de outro colega que respondeu de maneira muito parecida, pois o cansaço e a própria subjetividade humana podem influenciar a avaliação dos professores [Santos et al. 2016, Passero et al. 2016].

Visando ajudar nesses problemas, surge a correção assistida por computação, que tem como vantagens: formalização dos critérios de avaliação, feedback rápido (tanto para os alunos quanto para os professores) e consequentemente mais tempo disponível para o professor usar em suas outras atividades. Entretanto, avaliar respostas discursivas automaticamente não é uma tarefa simples para o computador e ainda é uma área de pesquisa na computação, com resultados como 70%, 80% e 90% de acurácia, dependendo de vários

fatores [Burrows et al. 2015]. Por isso, em um primeiro momento, essa tecnologia poderia ser aplicada em ambientes virtuais de aprendizagem, em um contexto de feedback rápido para o aluno, e não para sua avaliação final.

O uso de Ambientes Virtuais de Aprendizagem (AVAs) e a Educação à Distância (EAD) vem crescendo bastante no mundo e no Brasil. Em 2016, o Censo EAD.BR contabilizou 561.667 alunos matriculados em cursos regulares totalmente à distância. Além disso, mais quase 3 milhões de brasileiros realizam algum curso livre (não regulamentado) utilizando o EAD e AVAs para sua aprendizagem [ABED 2016]. Diante desses números, fica clara a necessidade de um suporte aos professores para a avaliação automática de exercícios discursivos, retirando o trabalho exaustivo e mecânico da avaliação feita atualmente pelo professor.

É nesse contexto que entra o sistema Auto-Avaliador Colaborativo e Inteligente de Respostas (Auto-Avaliador CIR), um ambiente online para o desenvolvimento das dinâmicas envolvendo questões e respostas em um contexto de aprendizado. Nele, professores podem se inscrever e criarem provas e questões a serem respondidas por estudantes. Por sua vez, estudantes terão acesso às provas e questões cadastradas pelos professores e poderão enviar suas respostas. Professores terão então acesso às respostas dos estudantes e poderão avaliá-las, fornecendo feedback ao aluno, que poderá ver suas notas em seu espaço no site.

O colaborativo da sigla vem do fato de que professores poderão colaborar entre si e com o sistema, principalmente de duas maneiras. A primeira é que um professor pode acrescentar respostas de referência e conceitos às questões de seus colegas, complementando e melhorando as questões. Além disso, professores poderão avaliar as respostas de quaisquer provas e questões do sistema, de forma a contribuir com o *feedback* do aluno.

O inteligente vem do fato de que o sistema, quando possuir respostas e suas respectivas avaliações, suficientes para uma questão, poderá avaliar automaticamente a resposta dos alunos subsequentes. Isso é possível devido a utilização de algoritmos de Aprendizado de Máquina, um dos ramos da Inteligência Artificial. Porém, para o correto funcionamento do sistema, muitas respostas e avaliações são necessárias, por isso quanto mais professores e alunos participarem colaborando melhor o sistema será, em termos de acurácia da nota gerada automaticamente.

2. Desenvolvimento

O processo de desenvolvimento do software foi iniciado com a concepção do relacionamento dos dados, em forma de um diagrama entidade-relacionamento. Após isso veio a decisão da linguagem/*framework* a ser utilizada para a programação, onde Python foi o escolhido.

A escolha de Python para programação web é de certa forma incomum visto que muitas aplicações utilizam PHP, Java ou Javascript. Porém, a escolha é justificada pelas bibliotecas utilizadas para o processamento de linguagem natural e aprendizado de máquina, respectivamente o NLTK [Bird and Loper 2004] e o Scikit-learn [Pedregosa et al. 2011]. Essas bibliotecas são umas das mais importantes para o desenvolvimento em suas áreas e portanto foram escolhidas para implementação do sistema de avaliação automática de respostas.

Visando criar um sistema unificado, foi usado Python para a aplicação web também, utilizando o *framework* Django¹ para um desenvolvimento mais rápido e focado. Para a interface foi utilizado o tradicional HTML e CSS, através do *framework* Bootstrap para agilizar e padronizar as telas, além de tornar o sistema responsivo para o uso em diversos dispositivos. A arquitetura utilizada pelo sistema, banco de dados *MySQL* e dispositivos de acesso pode ser visto na Figura 1.

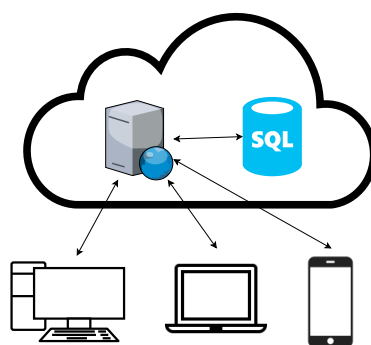


Figura 1. Arquitetura Geral do Sistema

O *framework* Django utiliza o conceito do padrão *Model-Template-View* (MTV), onde a modelagem dos dados fica à cargo da camada *model*, a interface do sistema é feita com *templates* e o *view* é a camada que cuida dos direcionamentos e controle geral da aplicação. Um modelo bem parecido com o muito utilizado *Model-View-Controller* (MVC). O Django utiliza também o mapeamento objeto-relacional, ficando a cargo do programador as classes de modelagem dos dados e o Django se encarrega da comunicação com o banco de dados.

2.1. Avaliação de Usabilidade

Com o objetivo de receber um feedback de usuários que utilizaram o sistema, nós aplicamos o teste de usabilidade *System Usability Scale* (SUS) [Brooke et al. 1996] com um professor e 29 alunos. O teste consiste em 10 perguntas que questionam o usuário sobre a usabilidade do sistema, incluindo sua facilidade de uso ou não, consistência entre outras questões. Ele utiliza a Escala Likert de 5 pontos em suas questões, pedindo ao respondente para responder à uma afirmação com o quanto ele concorda com a mesma, variando de “discordo completamente” até “concordo completamente”.

O Auto-Avaliador CIR obteve uma média de 75,1/100 no teste de usabilidade SUS, com as avaliações individuais variando de 42,5 à 97,5. É comumente aceito pela comunidade de usabilidade que a média para esse teste é de 68, portanto uma nota maior que essa média indica que o sistema possui uma boa usabilidade. Como o valor obtido pelo Auto-Avaliador CIR é apenas um pouco maior que a média, ainda há muito o que se melhorar, principalmente em relação ao seu design, como indicado em uma questão de texto livre incluída no questionário.

Além das questões do SUS e a que perguntava sobre o que pode ser melhorado no sistema, fizemos mais uma questão utilizando a Escala Likert, com a seguinte afirmação: “Eu acho que um sistema de avaliação automática de respostas discursivas é muito útil,

¹www.djangoproject.com/

especialmente em um contexto de curso à distância, através de um ambiente virtual de aprendizagem.”. Para essa afirmação, apenas um usuário discordou completamente, cinco foram neutros, quatro concordaram e 20 concordaram completamente. Utilizando o valor 1 para discordo completamente até 5 para concordo completamente, o Auto-Avaliador CIR obteve 4,4/5 para essa afirmação, indicando que os usuários majoritariamente concordam que a avaliação automática pode ser muito útil.

3. Apresentação do software

Como já exposto, o objetivo do Auto-Avaliador CIR² é fornecer um ambiente virtual para a realização de provas discursivas, a ser utilizado por professores e alunos, de forma totalmente gratuita³. O primeiro passo ao se cadastrar no sistema é escolher qual tipo de usuário será a sua conta, como visto na Figura 2. No caso da escolha como professor, é necessário que após o cadastro o usuário envie um e-mail para o administrador comprovando que é um professor, para evitar que qualquer pessoa possa se cadastrar como professor e avaliar respostas de alunos. Para o usuário estudante não há essa necessidade.



Figura 2. Escolha do tipo de usuário

A partir do cadastro, o sistema é dividido em dois, onde professores terão acesso à algumas ações e alunos à outras. Nas Figuras 3 e 6 é possível observar o menu principal de professores e de estudantes, com as ações mais relevantes do sistema.

3.1. Ações do Sistema

A seguir estão descritas as ações mais importantes que podem ser realizadas por usuários do sistema, dividido por usuários professores e estudantes.

Para **usuários professores**:

- **Cadastrar nova prova:** cadastra uma nova prova em branco com seu título e descrição.
- **Criar nova questão:** cadastra uma nova questão à uma prova, junto com possíveis respostas de referência esperadas pelo professor e conceitos envolvidos na questão (como palavras-chave).

²www.autoavaliadorcir.com

³youtu.be/nacBmterQsk

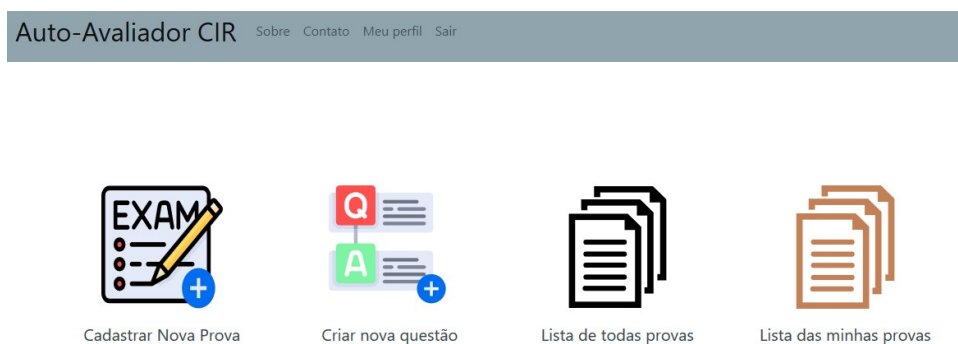


Figura 3. Menu principal do professor

- **Listar todas provas:** lista todas as provas cadastradas para visualização.
- **Listar minhas provas:** lista apenas as provas criadas pelo usuário.
- **Ver questões de uma prova:** mostra todas as questões de uma prova específica, como visto na Figura 4.

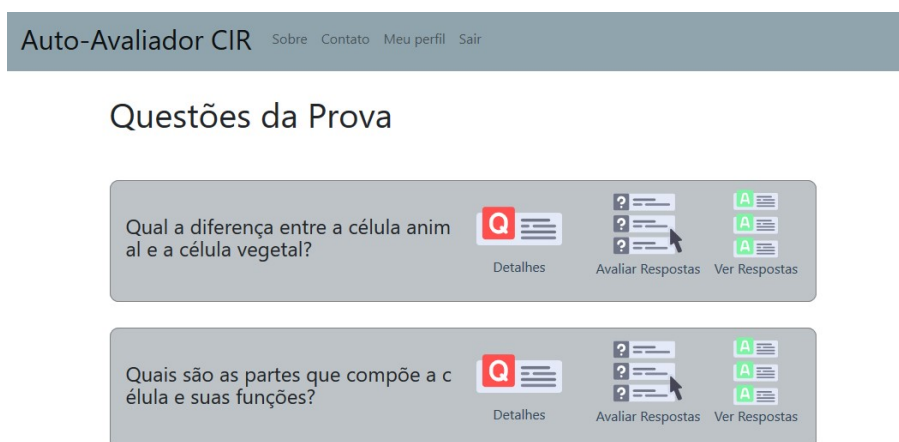


Figura 4. Lista de questões de uma prova

- **Detalhes de uma questão:** mostra todos os detalhes de uma questão, como seu enunciado, respostas de referência e conceitos cadastrados.
- **Ver respostas de uma questão:** mostra todas as respostas dos alunos para uma questão.
- **Avaliar respostas de uma questão:** apresenta uma resposta por vez de uma questão específica para o professor atribuir uma nota. Ele tem a opção de avaliar e terminar ou avaliar e continuar (avaliando a próxima resposta), como visto na Figura 5.

Auto-Avaliador CIR Sobre Contato Meu perfil Sair

Avaliar resposta

Falta avaliar mais **246** respostas além dessa.

Questão
Quais São As Partes Que Compõe A Célula E Suas Funções?

Resposta
Citoplasma (líquido que fica entre a membrana plasmática e núcleo), membrana plasmática (protege a célula) e núcleo (o que comanda a célula).

Nota
.....

[Avaliar e Terminar](#) [Avaliar e Continuar](#)

Conceitos envolvidos:

- membrana plasmática
- citoplasma
- núcleo

Respostas de Referência:

MEMBRANA PLASMÁTICA: O envelope externo da célula, que regula a troca de substâncias entre a célula e o meio externo.
CITOPLASMA: um material gelatinoso interno no qual estão mergulhados vários orgânulos menores. O citoplasma é o local em

Figura 5. Avaliação de respostas pelo professor

- **Adicionar resposta de referência:** adiciona uma resposta de referência à uma questão a escolha do professor. Diferentes professores podem colaborar com as provas cadastrando suas próprias respostas.
- **Adicionar conceito:** como a ação acima, realiza a adição de mais conceitos envolvidos no contexto da questão.

Para **usuários estudantes**, com menu principal representado na Figura 6:

Auto-Avaliador CIR Sobre Contato Meu perfil Sair

Ver todas questões ainda não respondidas por mim

Lista de provas

Lista das minhas respostas

Busca por provas ou questões

Figura 6. Menu principal do aluno

- **Ver todas questões ainda não respondidas por mim:** mostra ao aluno todas as questões cadastradas no sistema, para que ele possa escolher qual responder, das que ele ainda não respondeu.
- **Lista de provas:** mostra todas as provas disponíveis.
- **Lista das minhas respostas:** mostra uma lista com todas as respostas já feitas pelo usuário, com o enunciado e sua resposta, junto com a nota dada pelo professor que criou a questão, média da nota de outros professores e a nota dada pelo avaliador automático.
- **Busca por provas ou questões:** mostra um campo para buscar por provas de um professor específico, utilizando seu nome ou seu ID (o ID fica disponível ao professor em seu perfil).
- **Responder questão:** responde apenas uma questão isolada, da lista de todas as questões.

- **Responder questões:** responde uma prova completa, com todas as questões que a mesma possui, como visto na Figura 7.



Auto-Avaliador CIR Sobre Contato Meu perfil Sair

Prova Prova de Biologia

Uma prova com questões de biologia.

Qual a diferença entre a célula animal e a célula vegetal?

Quais são as partes que compõe a célula e suas funções?

Figura 7. Parte de uma prova de biologia com algumas questões

4. Considerações Finais

Este trabalho apresentou o Auto-Avaliador CIR, uma ferramenta colaborativa e inteligente para correção automática de respostas discursivas. Ele foi criado com o objetivo de auxiliar professores no processo de avaliação de seus alunos, pois ao realizar o processo automaticamente libera o professor para se dedicar a suas outras atividades, como a preparação de aulas.

Em um primeiro momento, seu uso poderia se dar no contexto de avaliações em que um rápido feedback é necessário, tanto para alunos quanto para os professores. Sua aplicação poderia ocorrer através de cursos online, que já utilizam a estrutura da web para realizar suas atividades. Já acostumados ao contexto virtual, alunos poderiam receber feedback através da resolução de questões discursivas e descobrir o que já está suficiente e o que é necessário estudar mais, de maneira automatizada.

O sistema já foi utilizado por cerca de 245 estudantes do ensino fundamental e atualmente conta com a correção automática apenas para 15 questões de biologia. Porém, seu uso por mais instituições, professores e alunos poderia rapidamente transformá-lo em um produto com avaliações automáticas para diversas questões de várias áreas do conhecimento.

O Auto-Avaliador CIR ainda está em fase inicial de seu processo como produto, mas poderia ser futuramente trabalhado como um *web service*, para permitir fácil integração com outros sistemas de ambientes virtuais de aprendizagem. Porém, já atualmente, um professor poderia começar com sua utilização e quando com um número razoável de respostas por questão fosse atingido, poderia aproveitar de sua avaliação automática, onde o uso contínuo do sistema melhoraria constantemente seu desempenho.

Referências

- ABED (2016). *Censo EAD Brasil 2016 - Relatório Analítico de Aprendizagem a Distância no Brasil*.
- Bird, S. and Loper, E. (2004). Nltk: the natural language toolkit. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, page 31. Association for Computational Linguistics.
- Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.
- Burrows, S., Gurevych, I., and Stein, B. (2015). The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education*, pages 60–117.
- Jacomini, M. A. and Penna, M. G. d. O. (2016). Carreira docente e valorização do magistério: condições de trabalho e desenvolvimento profissional. *Pro.posições*, 27(2):177–202.
- Nascimento, M. d. G. C. d. A. and Santos, J. V. (2015). Sessão Especial 05 - Políticas educacionais e currículo: interfaces na educação infantil e ensino fundamental 37^a Reunião Nacional da ANPED – 04 a 08 de outubro de 2015, UFSC – Florianópolis.
- Passero, G., Haendchen Filho, A., and Dazzi, R. (2016). Avaliação do uso de métodos baseados em lsa e wordnet para correção de questões discursivas. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 27, page 1136.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Santos, J. C. A. d. et al. (2016). Avaliação automática de questões discursivas usando lsa. *Universidade Federal do Pará*.
- Silva, G. L. F. and Rosso, A. J. (2008). As Condições Do Trabalho Docente Dos Professores Das Escolas Públicas De Ponta Grossa – Pr. *VII Congresso Nacional de Educação da PUCPR - EDUCERE e no III Congresso Ibero-Americano sobre Violências nas Escolas - CIAVE*, pages 495–536.