

Geração Automática de Questões a Partir do Reconhecimento de Entidades Nomeadas em Textos Didáticos

Juliana P. C. Pirovani¹, Marcos A. Spalenza¹, Elias de Oliveira¹

¹Programa de Pós-Graduação em Informática - Universidade Federal do Espírito Santo (UFES), 29060-970 - Vitória - ES - Brasil

juliana.campos@ufes.br, elias@lcad.inf.ufes.br

Abstract. *Learning evaluation is fundamental for the detection of methodological problems of the teaching practices. An effective assessment of teaching should address much of the content of the discipline and evaluate the comprehension of this content by the learners. In this work, we present a strategy that aims at automatic question generation from named entities identified in teaching texts. Therefore, we are proposing a tool that will save time of the teacher. In our experiments, free-text and fill-in-the-blank questions were generated from 10 types of named entities identified in a History textbook. We found 7195 named entities and we generate 6917 questions.*

Resumo. *A avaliação de aprendizagem é fundamental para a detecção dos problemas metodológicos do ensino. Uma verificação eficiente do ensino deve abordar boa parte do conteúdo da disciplina e avaliar a compreensão desse conteúdo pelos aprendizes. Neste trabalho apresentamos uma estratégia que visa a geração automática de questões a partir das entidades nomeadas identificadas em textos didáticos. Portanto, estamos propondo uma ferramenta que poupará tempo do professor. Em nossos experimentos, questões discursivas e questões de preenchimento foram geradas a partir de 10 tipos de entidades nomeadas identificadas em um livro didático de História. Encontramos 7195 entidades nomeadas e geramos 6917 questões.*

1. Introdução

A avaliação de aprendizagem é uma ferramenta fundamental para enfatizar os problemas do ensino. O objetivo da avaliação é apoiar e incentivar o processo de construção do conhecimento do estudante, seja de forma presencial ou à distância [Behar et al. 2010]. Uma verificação eficiente do ensino deve abordar boa parte do conteúdo da disciplina, onde estudantes são avaliados segundo a leitura e compreensão do texto. A proposta deste trabalho é auxiliar o professor na geração de questões para compor exercícios e avaliações a partir de conteúdos chave do texto, observando as relações entre entidades como sugerido por [Moraes 2003].

De acordo com [Rus et al. 2008], Geração de Questões (GQ) é a geração automática de questões a partir de textos escritos em linguagem natural, dados brutos e bases de conhecimento. Essa é uma tarefa importante que pode ser utilizada pelo professor, possibilitando diminuir o tempo e esforço necessários durante sua execução manual.

Nos últimos anos, muitas comunidades de pesquisa têm se interessado na geração automática de questões para propósitos educacionais [Le et al. 2014]. A GQ pode ser

utilizada em sistemas tutores inteligentes para dar *feedback* rápido ao aluno sobre a compreensão de um conteúdo específico; em sistemas de resposta automática a perguntas para prever as perguntas que podem ser feitas; e em ambientes educacionais em geral.

Este artigo apresenta uma abordagem para geração automática de questões com o intuito de avaliar o conhecimento adquirido pelo leitor do texto. Ou seja, o objetivo deste trabalho foi desenvolver uma aplicação educacional de GQ. A abordagem tem como diferencial o uso de dez tipos de Entidades Nomeadas (EN), como nomes de pessoas, locais e acontecimentos, identificadas no texto para GQ. A idéia por trás dessa abordagem é que EN aparecem com frequência em textos e são uma fonte de informação essencial para compreensão do seu conteúdo.

Dois tipos de questões foram geradas automaticamente a partir do Reconhecimento de Entidades Nomeadas (REN): questões de preenchimento (*fill-in-the-blank*) e questões discursivas (*free-text*) de resposta curta. Para isso, utilizamos as sentenças extraídas de um livro didático de História do Ensino Médio. O REN foi realizado por um sistema baseado em Campos Aleatórios Condicionais (*Conditional Random Fields - CRF*) e Gramáticas Locais (*Local Grammars - LG*).

A principal contribuição deste trabalho é que ele apresenta uma ferramenta capaz de produzir automaticamente, e em boa quantidade, questões sobre um determinado texto. O professor pode revisar e selecionar questões para compor uma avaliação posteriormente. Bancos de questões como os gerados por essa ferramenta são exigidos por alguns sistemas para criação automática de avaliações como os apresentados por [Tesseroli et al. 2016] e [Junior and Costa 2009].

Este artigo está organizado em 5 seções. A Seção 2 apresenta uma revisão de literatura. A Seção 3 apresenta a metodologia usada para identificação das EN e GQ. Os resultados são apresentados na Seção 4 e a Seção 5 apresenta as conclusões e trabalhos futuros.

2. Revisão de Literatura

Existem várias técnicas de Geração de Questões. Porém, grande parte dos sistemas de GQ são compostos quase que exclusivamente por regras escritas manualmente. Essas regras transformam representações linguísticas extraídas em etapas de pré-processamento em questões [Heilman 2011].

[Kalady et al. 2010] usa manipulação da árvore sintática das sentenças, REN e regras de transformação codificadas manualmente para GQ sobre fatos e definições (*factoid and definitional questions*). Stanford NER foi usado para REN e os autores afirmam que para gerar outros tipos de questões precisariam reconhecer outros tipos de entidades, além de pessoa, lugar e organização. Uma abordagem semelhante é apresentada em [Heilman 2011]. A principal diferença foi o uso de um classificador estatístico de perguntas que aproximadamente dobrou a taxa de aceitabilidade das melhores questões (*top-ranked*) apresentadas. Também foi apresentado um estudo de caso com professores, demonstrando o potencial de uso da GQ e seu impacto na educação.

Já [Eom et al. 2016] propõe uma abordagem para GQ do tipo verdadeiro ou falso a partir de um texto de entrada. Várias sentenças falsas são geradas a partir de uma sentença do texto substituindo palavras-chave por antônimos ou palavras erradas ou adicionando

um advérbio de negação ao verbo. A questão é gerada realocando a sentença verdadeira e as falsas randomicamente.

Algumas formas alternativas para a GQ são apresentadas em [Al-Yahya 2011] e [Zheng et al. 2011]. O sistema OntoQue [Al-Yahya 2011] usa ontologias de domínio para gerar itens de avaliação de três tipos: múltipla escolha, verdadeiro ou falso e de preenchimento. O conhecimento sobre entidades do domínio como classes, propriedades e indivíduos é usado para gerar as questões. O sistema K2Q [Zheng et al. 2011] gera questões a partir de palavras-chave de entrada, com o objetivo de auxiliar o usuário a gerar questões bem formuladas para sistemas de perguntas e respostas. Os autores usaram bases de dados de perguntas para induzir *templates* de questões que foram usados posteriormente para gerar novas questões com mesmo padrão.

Todos os trabalhos citados anteriormente visam a GQ a partir de textos em Inglês. Vários outros trabalhos para o Inglês são apresentados em [Le et al. 2014]. Porém, ainda existem poucos trabalhos para GQ a partir de textos escritos em Português. Alguns deles são [Amancio 2011], [Diéguez et al. 2011] e [Correia et al. 2012].

[Amancio 2011] aborda a GQ como um método original de elaboração textual. Para gerar as perguntas foi usado um corpus simplificado manualmente que transformou sentenças de voz ativa para voz passiva, reduziu o número de sentenças com aposto e dividiu sentenças coordenadas e subordinadas em duas ou mais com o objetivo de diminuir a complexidade sintática das sentenças e os erros de *parser*. [Diéguez et al. 2011] usou raciocínio baseado em casos (*Case-based reasoning - CBR*) para determinar quais sentenças poderiam ser usadas por um módulo de GQ. A idéia do CBR é verificar se existe uma resposta com a mesma estrutura em uma base de conhecimento e fazer uma pergunta igual à que foi feita para essa resposta.

[Correia et al. 2012] propõe o uso de uma técnica de classificação supervisionada, *Support Vector Machine - SVM*, para definir se uma dada sentença pode ser usada como base para questões de preenchimento. Algumas características usadas para gerar o modelo SVM foram informações sintáticas e EN. As questões são geradas no contexto do REAP.PT (*READER-specific Practice for Portuguese*), um sistema de ensino de vocabulário do Português Europeu.

Como a GQ em Português ainda é pouco estudada e continua sendo um desafio, este trabalho apresenta um sistema de GQ discursivas e de preenchimento em textos escritos em Português. Visto que os sistemas de GQ que usam EN normalmente reconhecem as entidades mais estudadas (pessoa, lugar e organização), o sistema proposto reconhece 10 tipos de EN e usa essas entidades como palavras-chave para a geração automática de questões.

3. Metodologia

Um livro didático de História¹ em domínio público foi utilizado para GQ neste trabalho. O livro é recomendado para o Ensino Médio e aborda assuntos como: “Trabalho Escravo e Trabalho Livre”, “Urbanização e Industrialização”, “O Estado e as Relações de Poder” e

¹Livro didático de História da Secretaria de Educação do Estado do Paraná (2006). Distribuição livre e gratuita. Disponível em: http://www.educadores.diaadia.pr.gov.br/arquivos/File/livro_didatico/historia.pdf Acesso em: 11/04/2017

“Movimentos Sociais, Políticos e Culturais”. Esse livro de História foi escolhido para os experimentos pois contém muitas EN como nomes de pessoas, locais e eventos temporais.

O texto foi extraído a partir do pdf do livro usando a ferramenta Apache Tika². Cada capítulo do livro foi segmentado em sentenças utilizando a ferramenta Unitex³.

Na sequência, foi realizado o REN em cada arquivo contendo uma sentença do livro. Foram reconhecidas as 10 categorias de EN do HAREM [Mota and Santos 2008], um evento de avaliação conjunta em REN para o Português. As 10 categorias são: pessoa, local, organização, valor, tempo, acontecimento, abstração, obra, coisa e outro. A saída do processo de REN é o conjunto de sentenças com as marcações de EN.

Um filtro foi aplicado após o REN para remover algumas sentenças que não são consideradas boas para a geração de questões. Alguns exemplos de sentenças removidas por esse filtro são: sentenças que terminam com o caractere “?” pois já são perguntas; sentenças compostas somente pela EN pois não apresentam contexto para geração de perguntas e sentenças que começam com palavras como “Fonte:”, “Adaptado de”, “Disponível em”, etc, que contém referências e não contém informações relevantes.

As questões são geradas a partir das sentenças que tiveram pelo menos uma EN reconhecida. As EN reconhecidas foram usadas para determinar onde seriam incluídos espaços em branco para gerar questões de preenchimento (*fill-in-the-blank*). Já os tipos das EN foram usados para determinar o pronome interrogativo para gerar questões discursivas (*free-text*).

A Figura 1 apresenta uma visão geral da metodologia utilizada.

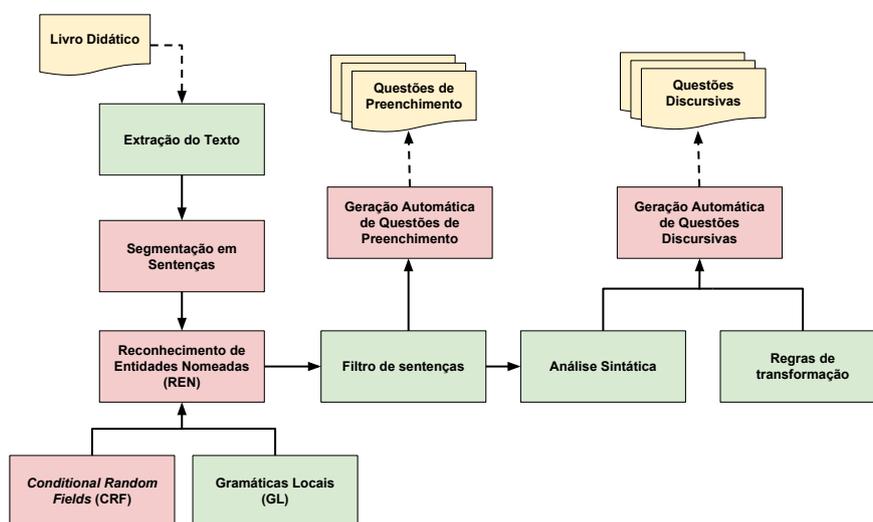


Figura 1. Visão geral da metodologia utilizada

3.1. Reconhecimento de Entidades Nomeadas

O Reconhecimento de Entidades Nomeadas (REN) foi realizado combinando a rotulação obtida por Conditional Random Fields (CRF) e a classificação obtida através de

²<https://tika.apache.org/>

³<http://unitexgramlab.org/>

Gramáticas Locais (GL). CRF é uma técnica de aprendizado de máquina que vem sendo utilizada com sucesso em diversas atividades de Processamento de Linguagem Natural (PLN) e GL são uma forma de representar regras contextuais onde EN podem aparecer. A classificação obtida pela GL é enviada como característica para aprendizado do modelo de predição CRF. Ou seja, a classificação atribuída pela GL pode ser vista como uma sugestão para o CRF que atribui o rótulo final das EN.

O modelo CRF foi treinado utilizando os corpus do HAREM⁴ (Primeiro HAREM, Mini HAREM e Segundo HAREM). O padrão do HAREM foi utilizado para marcação de EN. Assim, após o REN, cada sentença do texto possui as marcações de EN entre as tags e contendo a categoria da EN como no exemplo a seguir:

```
Os sucessores de <EM ID="Cap12-129"
CATEG="PESSOA"> Diocleciano </EM> continuaram a manter
o sistema de absolutismo.
```

3.2. Geração de Questões de Preenchimento

As questões de preenchimento são geradas para sentenças que possuem mais de uma EN para tentar garantir a existência de contexto suficiente para inferir a resposta. Foi gerada uma questão de preenchimento para cada EN identificada em uma sentença substituindo a EN por espaços em branco. O aluno deve identificar qual palavra ou sequência de palavras preenche corretamente esses espaços.

Um exemplo da geração de questões de preenchimento é apresentado na Tabela 1.

Tabela 1. Exemplo da geração de questões de preenchimento

Sentença	Em 1570, D. Sebastião proibiu a escravidão indígena, a não ser daqueles hostis ou antropófagos.
Sentença após REN	Em <EM ID="Cap4-291" CATEG="TEMPO"> 1570 , <EM ID="Cap4-292" CATEG="PESSOA">D. Sebastião proibiu a escravidão indígena , a não ser daqueles hostis ou antropófagos .
Questões geradas	Em _____, D. Sebastião proibiu a escravidão indígena, a não ser daqueles hostis ou antropófagos. Em 1570 , _____ proibiu a escravidão indígena, a não ser daqueles hostis ou antropófagos.

3.3. Geração de Questões Discursivas

A abordagem baseada em transformação, conforme citada em [Le et al. 2014], foi utilizada para gerar as questões discursivas. Vinte regras de transformação foram descritas e implementadas para transformar representações linguísticas das sentenças em questões. A representação linguística utilizada neste trabalho foi a árvore sintática das sentenças, extraída pelo analisador sintático (*parser*) do CoGrOO [Silva 2014], com a informação das EN reconhecidas.

O tipo de EN reconhecida indica o pronome interrogativo que será usado para formular a questão. Se a EN é do tipo pessoa, o pronome usado é *quem*; se é do tipo local, o pronome usado é *onde*; e assim por diante.

⁴Linguatca (2017). <http://www.linguatca.pt/HAREM/>. Acesso em: 11/04/2017.

A partir da árvore sintática é possível identificar os verbos que serão utilizados para fazer perguntas. O verbo deve ser convertido para uma representação correta. Por exemplo, para a sentença ``Os Stuarts retornaram ao poder'', o verbo deve ser convertido para a terceira pessoa do singular, pois a pergunta correta é ``Quem retornou ao poder''. Os dicionários do Unitex foram usados para conjugar o verbo na forma correta.

As regras de transformação são apresentadas da seguinte forma:

Estrutura da sentença → Estrutura da questão

Toda sentença que possuir a estrutura apresentada à esquerda da seta será usada para gerar uma questão com a estrutura apresentada à direita da seta.

O passo-a-passo da construção de uma questão discursiva é apresentado na Tabela 2. O CoGrOO usa o rótulo NP (*noun phrase*) para representar frases nominais, VP (*verbal phrase*) para frases verbais, PP (*prepositional phrases*) para frases preposicionadas e ADVP (*adverbial phrase*) para frases adverbiais.

Tabela 2. Exemplo da geração de questões discursivas

Sentença	A Revolução Francesa trouxe significativos avanços no tratamento da questão da tortura.
Sentença após REN	A <EM ID="Cap14-297" CATEG="ACONTECIMENTO" > Revolução Francesa trouxe significativos avanços no tratamento da questão da tortura.
Sentença após análise sintática	NP[A Revolução Francesa] VP[trouxe] NP[significativos avanços] PP[em o] NP[tratamento] PP[de a] NP[questão] PP[de a] NP[tortura]
Análise sintática com REN	NP[A ACONTECIMENTO:Revolução Francesa] VP[trouxe] NP[significativos avanços] PP[em o] NP[tratamento] PP[de a] NP[questão] PP[de a] NP[tortura]
Regra usada na GQ	<NP <ACONTECIMENTO>> <VP> ... → Qual acontecimento <VP> ...?
Questão gerada	Qual acontecimento trouxe significativos avanços no tratamento da questão da tortura?

A regra usada na GQ da Tabela 2 indica que a sentença possui a seguinte estrutura: inicia com uma frase nominal que contém uma EN do tipo Acontecimento (<NP <ACONTECIMENTO>>); é seguida por uma frase verbal (<VP>) e, na sequência, pode ter quaisquer rótulos (...). A questão será gerada usando a estrutura: inicia com Qual acontecimento; é seguida pela frase verbal (<VP>) que aparece na sentença após a frase nominal que contém a EN e o restante da frase (...).

Ao descrever as regras de transformação, consideramos quatro situações para gerar as questões:

1. a EN aparece como sujeito da sentença e executa uma ação indicada por um verbo específico (afirmou, proferiu, conforme, etc). Ex: A sentença Foi neste contexto que Cromwell proferiu esta frase: ``Não há outro modo de se lidar com estes homens a não ser partindo-os em pedaços'' gera a pergunta Quem proferiu esta frase: Não há outro modo de se lidar com estes homens a não ser partindo-os em pedaços?

2. o posicionamento do verbo em relação a EN é analisado para geração da pergunta. Ex: A sentença A medida adotada pela Companhia evitou conflitos entre os colonos antigos e os recém-chegados gera a pergunta Qual organização evitou conflitos entre os colonos antigos e os recém - chegados?
3. a EN identificada é substituída pelo pronome interrogativo adequado. Ex: A sentença O tamanho médio dos lotes coloniais, em geral, foi de 15 mil alqueires gera a pergunta O tamanho médio dos lotes coloniais, em geral, foi de quantos alqueires?
4. preposições e outras palavras próximas (após, antes, durante, etc) a EN indicam relações específicas com a sentença e são usadas na pergunta. Ex: A sentença O aumento da exportação da erva mate para os países platinos, se deu após a Guerra do Paraguai gera a pergunta O que aconteceu após a Guerra do Paraguai?

3.4. Avaliação

Segundo [Kalady et al. 2010], não existe uma forma padrão para avaliar a saída de sistemas de GQ. Logo, devido a grande quantidade de questões geradas, consideramos uma amostra aleatória de 10% para avaliação neste trabalho. Essa amostra foi submetida ao juízo de 15 alunos de um curso de graduação de nossa instituição de ensino para avaliação.

Os alunos atribuíram um dos seguintes códigos a cada pergunta avaliada: *Código 1* – Incorreta: Se a pergunta gerada não tiver algum significado, não fizer sentido; *Código 2* – Parcialmente correta: Se a pergunta gerada tiver algum erro, como um erro gramatical; *Código 3* – Plenamente correta: Se a pergunta é clara, não apresenta nenhum erro.

Cada pergunta foi considerada *incorreta*, *parcialmente correta* ou *plenamente correta* quando pelo menos dois avaliadores concordaram nesse sentido, ou seja, atribuíram o mesmo código a ela. Cada questão da amostra foi avaliada até obter a concordância em 5% das questões geradas.

4. Resultados e Discussão

Ao realizar o REN no livro didático de História foram identificadas 7195 EN. Essas EN foram identificadas em 3390 sentenças das 7012 existentes no livro. Após aplicar o filtro para remover as sentenças ruins, restaram 2106 sentenças contendo 5040 EN para a GQ.

A partir das EN identificadas foram geradas 6917 questões, sendo 4175 de preenchimento e 2742 discursivas. O número de questões discursivas geradas é menor pois essas questões não são geradas para todas as sentenças. Para gerar questões discursivas, a estrutura da sentença deve corresponder à estrutura apresentada do lado esquerdo de pelo menos uma regra de transformação. A Tabela 3 apresenta exemplos das questões geradas.

Foram avaliadas 700 questões selecionadas aleatoriamente. Das questões discursivas avaliadas por mais de um aluno, obtivemos concordância (dois alunos atribuíram o mesmo código) em 141 delas. Para as questões de preenchimento, obtivemos concordância em 211, alcançando os 5% das questões geradas.

A Tabela 4 apresenta a porcentagem de questões avaliadas pelos alunos com cada código (1, 2 ou 3) para os dois tipos de questões geradas.

Tabela 3. Exemplos de questões geradas automaticamente

Id	Sentença	Questão
1	A hora de Greenwich, adotada na Grã-Bretanha em 1848, chegou aos Estados Unidos em 1873, ao Japão em 1888 e ao Brasil em 1914.	A hora de Greenwich, adotada na _____ em 1848, chegou aos Estados Unidos em 1873, ao Japão em 1888 e ao Brasil em 1914.
2	O alerta consta do “Relatório Global sobre Aglomerações Humanas”, que resume o encontro (Habitat) promovido pela ONU em 1996 sobre o tema.	O alerta consta do “Relatório Global sobre Aglomerações Humanas”, que resume o encontro (Habitat) promovido pela _____ em 1996 sobre o tema.
3	Já na região Norte, Londrina, deu origem a outros municípios como: Cambé, Rolândia e Apucarana.	Já na região Norte, Londrina, deu origem a outros municípios como: Cambé, _____.
4	Rápida, a revolução urbana é um fenômeno da segunda metade deste século XX.	Quando a revolução urbana foi um fenômeno da segunda metade deste?
5	Mais de 20 mil pessoas viviam nesses pardieiros.	Quantos pessoas viviam nesses pardieiros?
6	Sua importância estava no fato de que o lar romano era protegido pela deusa Vesta.	Sua importância estava no fato de que o lar romano era protegido por quem?
7	Em 1993, os camponeses mexicanos se organizaram para a luta armada e realizaram os primeiros conflitos com o exército.	Quando os camponeses mexicanos se organizaram para a luta armada e realizaram os primeiros conflitos com o exército?

Tabela 4. Resultados

Código	Significado	% Preenchimento	% Discursivas
1	Incorreta	5.21%	24.82%
2	Parcialmente correta	3.31%	17.02%
3	Plenamente correta	91.48%	58.16%

Conforme Tabela 4, os avaliadores entenderam que 5.21% das questões de preenchimento e 24.82% das discursivas apresentadas não seriam inteligíveis por humanos (*Código 1*), como a questão 4 na Tabela 3. Por outro lado, 3.31% das questões de preenchimento e 17.02% das discursivas foram avaliadas como inteligíveis, porém com algum erro, talvez gramatical (*Código 2*). Exemplos de questões avaliadas com esse código são a 3 e a 5 na Tabela 3. Por fim, 91.48% das questões de preenchimento e 58.16% das discursivas foram consideradas *plenamente corretas*, prontas para uso pelo professor em suas atividades pedagógicas. As demais sentenças na Tabela 3 foram avaliadas com *Código 3*.

Observe que os pequenos erros das questões avaliadas com *Código 2* (questões 3 e 5) podem ser corrigidos posteriormente, até mesmo manualmente por um professor. Assim, considerando as questões de códigos 2 e 3 como inteligíveis, alcançamos a porcentagem de 94.79% de questões de preenchimento inteligíveis e 75.18% de questões discursivas inteligíveis.

Como esperado, o resultado obtido para as questões de preenchimento foi melhor do que o resultado obtido para as questões discursivas. Isso acontece porque a geração das questões de preenchimento depende dos processos apresentados na Figura 1 até a filtragem de sentenças, enquanto a geração de questões discursivas depende também da análise sintática e da transformação das sentenças, sendo todos esses processos suscetíveis a erros.

Por exemplo, na pergunta 3 da Tabela 3, Rolândia e Apucarana foi identificado como uma única EN e, por isso, na sentença gerada existe apenas um campo para preenchimento. Na pergunta 4, século XX foi identificado como EN do tipo Tempo ao invés de segunda metade deste século XX. Por isso, da segunda metade deste aparece no final da questão. Já na pergunta 5 da mesma Tabela, o erro de concordância foi inserido durante o processo de transformação da sentença porque nossa abordagem ainda não realiza a verificação de gênero da palavra.

Alguns outros problemas observados que contribuíram para a geração de perguntas consideradas incorretas foram: informações extras nas sentenças devido a erros no processo de extração do texto; sentenças incompletas devido a erros na segmentação do texto; e perguntas com pronome interrogativo inadequado por uma atribuição errada de tipo à EN.

É importante ressaltar que as perguntas são geradas a partir das sentenças originais do texto, o que significa que trabalhamos com muitas sentenças longas e complexas. Isso também contribuiu para geração de perguntas incorretas ou parcialmente corretas.

5. Conclusões

Neste trabalho propomos uma abordagem para a geração automática de questões centradas em 10 EN indentificadas em um livro didático de História. Do livro texto utilizado como base para os experimentos, identificamos 7195 entidades e geramos 6917 perguntas. Cada pergunta é gerada a partir da transformação da sentença, introdução de pronomes interrogativos e flexão do verbo ligado a entidade.

Para validar a qualidade das perguntas geradas pela nossa abordagem, alunos de um curso de graduação se tornaram juízes e avaliaram aquelas que fariam sentido ou não. Foram consideradas plenamente corretas para uso pelo professor 58.16% das questões discursivas e 91.48% das questões de preenchimento. Além disso, 17.02% das questões discursivas e 3.31% das questões de preenchimento foram consideradas parcialmente corretas, apresentando pequenos erros gramaticais que podem ser corrigidos posteriormente.

Como trabalhos futuros, pretendemos melhorar cada vez mais os processos envolvidos na geração de questões, como o REN, a análise sintática e as regras existentes. Novas regras também podem ser inseridas para gerar um número maior de questões discursivas. Além disso, pensamos em avaliar o sistema desenvolvido em outros textos didáticos e em textos disponíveis na internet.

Também pensamos em avançar nossas pesquisas na capacidade de fazer perguntas mais complexas das que apresentamos neste trabalho. Nossa aposta vai no sentido de utilizar caminhos mais longos, gerando perguntas cujas respostas exijam algum tipo de inferência.

Referências

Al-Yahya, M. (2011). OntoQue: A Question Generation Engine for Educational Assessment Based on Domain Ontologies. In *Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on*, pages 393–395. IEEE.

- Amancio, M. A. (2011). Elaboração textual via definição de entidades mencionadas e de perguntas relacionadas aos verbos em textos simplificados do português. Master's thesis, Universidade de São Paulo.
- Behar, P. A., Ribeiro, A. C. R., Longaray, A. N. C., Bernardi, M., and da Silva, K. K. A. (2010). AVALEAD: um objeto de aprendizagem sobre avaliação em Educação a Distância. In *Simpósio Brasileiro de Informática na Educação - SBIE 2010*, volume 1.
- Correia, R., Baptista, J., Eskenazi, M., and Mamede, N. (2012). Automatic generation of cloze question stems. In *International Conference on Computational Processing of the Portuguese Language*, pages 168–178. Springer.
- Diéguez, D., Rodrigues, R., and Gomes, P. (2011). Using CBR for Portuguese Question Generation. In *Proceedings of the 15th Portuguese Conference on Artificial Intelligence*, pages 328–341.
- Eom, J.-S., Lim, U.-K., and Lee, J.-Y. (2016). Automatic Generation System of True or False Questions. *Advanced Science and Technology Letters*, 135:46–49.
- Heilman, M. (2011). *Automatic factual question generation from text*. PhD thesis, Carnegie Mellon University.
- Junior, P. A. P. and Costa, H. A. X. (2009). SGSO—Um Sistema Baseado em Lógica Fuzzy para Geração e Correção de Simulados e Provas via Web. In *WIE XV Workshop Sobre Informática na Escola, Publisher: XXIX Congresso da Sociedade Brasileira de Computação - WIE 2009*, volume 1, pages 1881–1890.
- Kalady, S., Elikkottil, A., and Das, R. (2010). Natural Language Question Generation Using Syntax and Keywords. In *Proceedings of QG2010: The Third Workshop on Question Generation*, pages 1–10. questiongeneration.org.
- Le, N.-T., Kojiri, T., and Pinkwart, N. (2014). *Automatic Question Generation for Educational Applications – The State of Art*, pages 325–338. Springer International Publishing, Cham.
- Moraes, R. (2003). Uma Tempestade de Luz: a Compreensão Possibilitada pela Análise Textual Discursiva. *Ciência & Educação (Bauru)*, 9:191 – 211.
- Mota, C. and Santos, D. (2008). Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM.
- Rus, V., Cai, Z., and Graesser, A. (2008). Question Generation: Example of A Multi-year Evaluation Campaign. *Proc WS on the QGSTEC*.
- Silva, W. D. C. d. M. (2014). Aprimorando o Corretor Gramatical CoGrOO. Master's thesis, Instituto de Matemática e Estatística - Universidade de São Paulo (USP), São Paulo, SP.
- Tesseroli, R., Direne, A., Pimentel, A., Spinosa, E., and Melniski, L. (2016). Geração automática de avaliações utilizando o algoritmo ABC para definição da próxima questão. In *Anais dos Workshops do V Congresso Brasileiro de Informática na Educação - CBIE 2016*, volume 5, page 407.
- Zheng, Z., Si, X., Chang, E. Y., and Zhu, X. (2011). K2Q: Generating Natural Language Questions from Keywords with User Refinements. In *IJCNLP*, pages 947–955.