

Atividades Lúdicas baseadas em Expressões Regulares ou Gramáticas

Engenharia Gramatical

Jéssica Fernandes
Universidade do Minho
Email: a93318@alunos.uminho.pt

Vitor Lelis
Universidade do Minho
Email: pg54273@alunos.uminho.pt

1. Introdução e Contextualização

A engenharia de linguagens desempenha um papel crucial no desenvolvimento de software, influenciando diretamente a qualidade e eficiência do código produzido.

No contexto da engenharia de linguagens, surgem dois conceitos fundamentais:

- 1) A engenharia gramatical, que compreende a construção e análise de gramáticas formais, que descrevem a estrutura das linguagens, especificando como as construções sintáticas e semânticas devem ser organizadas.
- 2) As expressões regulares, que se relacionam diretamente com a engenharia gramatical, pois são frequentemente incorporadas em gramáticas formais, para especificar padrões mais simples dentro da sintaxe global da linguagem.

A sua compreensão é essencial, porém, por vezes, desafiante. Felizmente, existem diversas ferramentas interativas bem como atividades lúdicas, baseadas nestes conceitos, que procuram facilitar a sua aprendizagem de forma muito prática e eficiente.

Será apresentada, de seguida, uma compilação de ferramentas e atividades deste tipo, explicando e comparando os seus propósitos, e a forma como aplicam os conceitos mencionados.

2. Ferramentas Analisadas

2.1. Expressões Regulares

2.1.1. Regulex. O Regulex [1] é uma ferramenta visual que permite aos utilizadores criar e visualizar expressões regulares de maneira gráfica. O utilizador insere uma expressão e a ferramenta produz a representação visual e sequencial da expressão, elemento a elemento.

Esta ferramenta é ótima para a compreensão dos padrões de correspondência pois representa cada parte de forma detalhada, e destaca grupos de captura na expressão, permitindo detetar como os diferentes elementos se combinam, e qual o propósito de cada um.



Figure 1. Exemplo 1 - Regulex

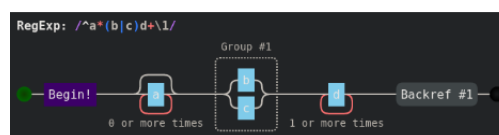


Figure 2. Exemplo 2 - Regulex

Esta forma de visualização é muito acessível para principiantes, e principalmente para utilizadores que preferem aprender de uma forma visual.

Quanto mais complexas as expressões se tornarem, mais difícil será lê-las e entendê-las apenas com o suporte textual. Com esta ferramenta, é possível visualizar a forma como se representa cada símbolo individualmente, ou associado a outros, criando as nossas próprias sequências que nos permitem aprender de forma personalizada e sem restrições.

O Regulex é um projeto *open source*, e por isso possui o seu repositório público no Github [2].

2.1.2. RegexOne. O RegexOne [3] consiste num conjunto de exercícios interativos nos quais o objetivo é criar uma expressão regular, que identifique uma lista de palavras simultaneamente.

Cada nível possui uma contextualização e descrição detalhadas do que deve ser feito, bem como uma breve explicação do conceito que deve ser utilizado e com que propósito.

A dificuldade aumenta progressivamente com o decorrer dos exercícios e a introdução de conceitos mais complexos. Em alguns níveis aparecem algumas palavras que devem ser ignoradas, o que aumenta a dificuldade também.

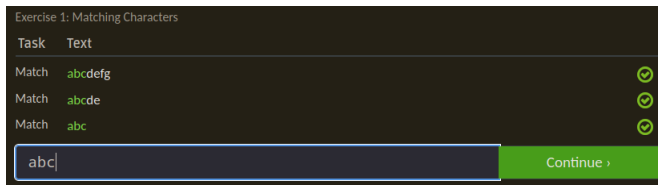


Figure 3. Exemplo 1 - RegexOne

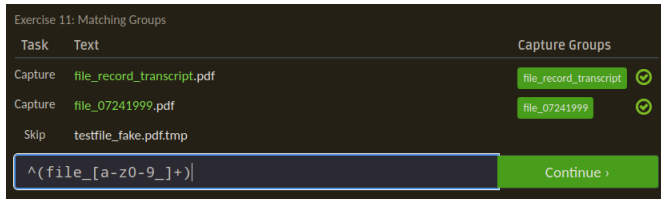


Figure 4. Exemplo 2 - RegexOne

2.1.3. Regex Golf. Podemos considerar o Regex Golf [4] um jogo simples, parecido ao anteriormente mencionado, sendo a principal diferença, a falta de orientação e explicação dos conceitos.

Neste jogo, além de possuirmos uma lista de palavras extensa que devem ser identificadas, é introduzida uma lista de palavras, igualmente longa, que devem ser ignoradas pela expressão regular.

O desafio é identificar o padrão a ser identificado de um lado, e ignorado do outro, e criar uma expressão regular que simultaneamente seja compatível com um dos conjuntos, e incompatível com o outro.

A dificuldade aumenta progressivamente com o decorrer dos níveis, obrigando o jogador a conhecer mais símbolos, abrindo um leque de novas opções, e refinando o seu domínio das expressões.

Um exemplo:

Neste nível foi utilizado o símbolo \$ para indicar que a palavra poderia conter o segmento 'ick' antecedido por qualquer coisa, porém, não seguido por nada. No nível seguinte, o raciocínio é o mesmo, porém, é exigido que não se use o símbolo \$.

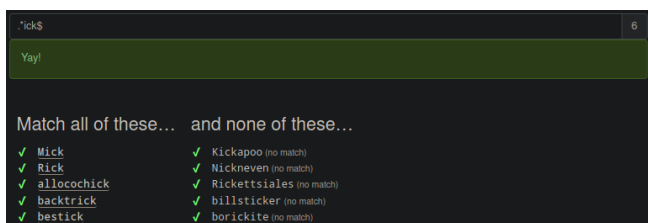


Figure 5. Uso do \$ no Regex Golf

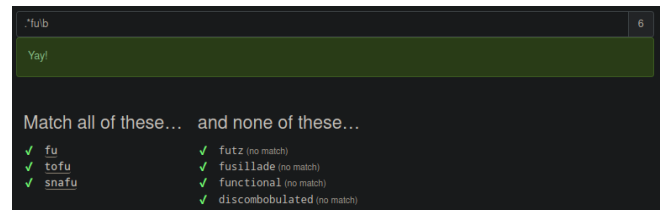


Figure 6. Alternativa ao uso do \$ no Regex Golf

2.1.4. Regex Crossword. O Regex Crossword [5] consiste num jogo muito interessante de palavras cruzadas com expressões regulares.

O objetivo é preencher cada quadrado com um caracter que satisfaça as restrições da linha e coluna onde se encontra.

Em semelhança aos anteriores, a dificuldade aumenta progressivamente com o decorrer dos níveis, introduzindo mais símbolos e conceitos.

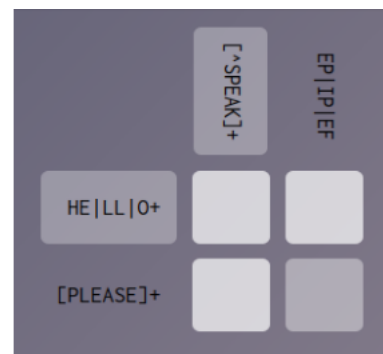


Figure 7. Nível Fácil do Regex Crossword



Figure 8. Nível Difícil do Regex Crossword

2.1.5. Comparações. Em termos de aprendizagem, o **Reg-ulex** e o **RegexOne** são ótimos pontos de partida para

embarcar na aventura de expressões regulares pois possuem explicações dos conceitos. Particularmente, o **RegexOne**, como já mencionado, providencia o conteúdo escrito, sobre o qual incidem os exercícios, à medida que são apresentados.

Já o **Regex** não possui descrições detalhadas, porém, apresenta a explicação das expressões, à medida que são inseridas, de forma visual. Isto permite uma compreensão eficiente do propósito de cada elemento.

Em suma, ambos são bons a nível de exposição dos conceitos, e possibilitam a introdução à aprendizagem, de formas alternativas, podendo até serem complementares.

O **Regex Golf** e o **Regex Crossword** são perfeitos para quem já possui bases e pretende praticar os conceitos, pois não possuem tanto suporte em termos de conteúdo.

O **RegexGolf** é conceitualmente semelhante ao **RegexOne**, mas o nível de dificuldade é mais elevado pelo acréscimo de uma nova lista de palavras, e pelo facto de não ser providenciada nenhuma informação aquando da resolução do desafio, sendo, puramente prático.

O **Regex Crossword** torna-se mais desafiante do que os anteriores, pois, as expressões que devem ser identificadas são apresentadas de forma mais abstrata, como regras e não como palavras. Adicionalmente, torna-se mais complexo, atingindo um nível de dificuldade mais elevado no final e possuindo um leque mais variado de níveis. É, sem dúvida, o mais completo a nível prático.

2.2. Gramáticas

2.2.1. Hitchhiker's Guide to the Galaxy. Baseado no livro de mesmo nome escrito por Douglas Adams, este jogo de ficção Interativa foi originalmente lançado em 1984 pelo autor e Steve Meretzky para Apple II, Macintosh, Commodore 64, CP/M, MS-DOS, Amiga e Atari ST. No entanto, para comemorar os 30 anos da obra, foi lançada uma nova versão para navegador [6], proporcionando um acesso mais fácil.

Assim como outros jogos do mesmo género, a mecânica do jogo envolve descrições de situações em que o jogador se encontra. O jogador deve então descrever as ações que deseja realizar, tornando-se um participante ativo que influencia o curso de toda a cena.



Figure 9. Cena do jogo Hitchhiker's Guide to the Galaxy

No que diz respeito à sua funcionalidade gramatical, o processamento dos comandos do jogador envolve a

tokenização das palavras para identificar termos-chave que podem gerar cenários possíveis dentro da situação atual.

2.2.2. CodeCombat. A plataforma CodeCombat [7] é uma ferramenta inovadora de aprendizagem de programação, que se utiliza de jogos como meio para apresentar desafios e missões aos usuários. A peculiaridade reside no facto de que os comandos necessários para completar esses objetivos são expressos por meio da escrita de código.



Figure 10. Cenário com Objetivo - CodeCombat

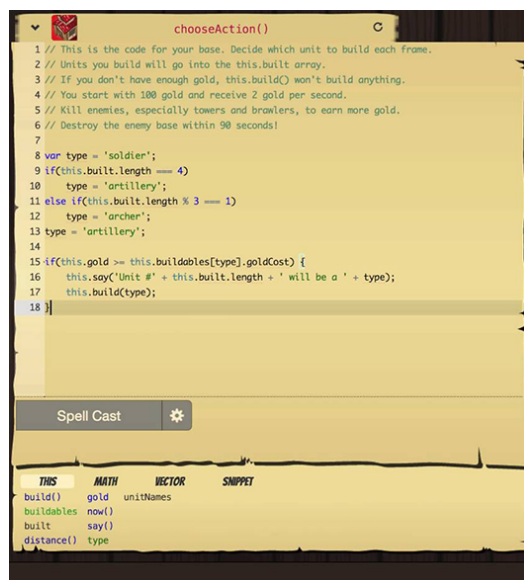


Figure 11. Linha de comandos - CodeCombat

A Figura 10 apresenta o desafio atual em que o jogador se encontra, determinando as metas necessárias para avançar no seu progresso. Esses desafios podem abranger diversas tarefas, tais como:

- 1) Navegar de um ponto a outro no cenário;
- 2) Derrotar um ou mais inimigos estrategicamente posicionados;
- 3) Esquivar-se de armadilhas engenhosas;
- 4) E outras situações igualmente variadas.

Simultaneamente, a Figura 11 exibe a linha de comandos onde o jogador deve formular as ações necessárias para atingir os seus objetivos. Como já mencionado, essas ações são expressas por meio da escrita em linguagens de programação como Python, JavaScript, Java, entre outras.

Embora a ênfase desta plataforma não esteja na aprendizagem de conceitos gramaticais, a aplicação desses conceitos é indispensável para o seu funcionamento. Nesse ambiente, onde a semântica e a gramática da linguagem de programação específica são cruciais para o progresso do jogador, a existência de ferramentas capazes de analisar, corrigir e processar o código escrito é fundamental para assegurar o bom funcionamento da plataforma.

O CodeCombat também se trata de uma organização com repositórios *open source* [8].

2.2.3. ANTLRWorks. O ANTLRWorks [9] é um Ambiente de Desenvolvimento Integrado (IDE) projetado para facilitar o uso do ANTLR (ANother Tool for Language Recognition), oferecendo uma interface gráfica que auxilia na análise léxica e sintática de linguagens.

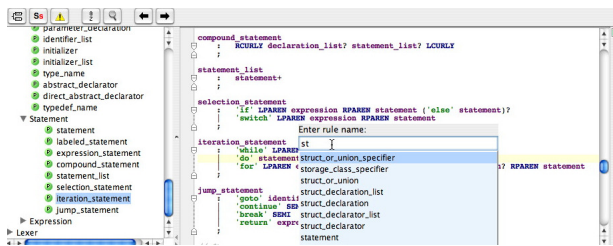


Figure 12. ANTLRWorks

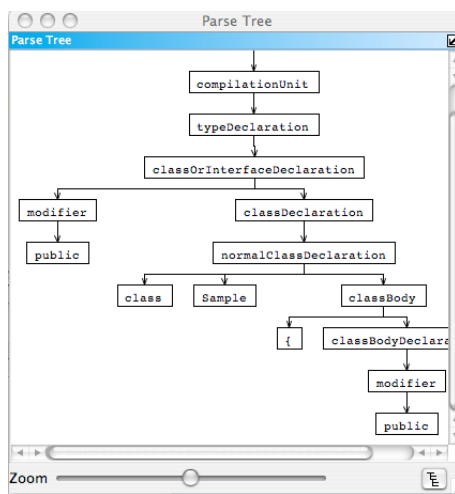


Figure 13. ANTLRWorks Tree Parser

Em termos de funcionalidades, este IDE desempenha um papel crucial na criação e depuração de gramáticas complexas, proporcionando recursos como realce de sintaxe, e possibilitando uma visualização mais clara da estrutura da gramática. Além disso, oferece a capacidade de gerar e examinar *parser trees*, que facilitam a identificação dos termos, de acordo com a gramática proposta. O ANTLRWorks também disponibiliza diversas outras funcionalidades que se integram naturalmente no fluxo de trabalho cotidiano com o ANTLR.

2.2.4. Comparações. Enfrentamos uma certa dificuldade em encontrar opções lúdicas na procura e análise de ferramentas que utilizam gramáticas. Embora tenhamos identificado aplicações com essa proposta, como o **ANTLRWorks**, observamos que, no requisito lúdico, não proporcionava a experiência desejada. Perante este desafio, decidimos direcionar a nossa pesquisa para uma abordagem diferente, procurando ferramentas lúdicas que incorporam conceitos de análise semântica e léxica no seu funcionamento.

Levando este ponto em consideração, as comparações entre as ferramentas escolhidas serão feitas levando em consideração os requisitos: funcionalidade baseada em gramáticas e potencial lúdico.

No primeiro requisito, a ferramenta vencedora foi o **ANTLRWorks**, devido ao seu ambiente dedicado ao estudo e desenvolvimento de gramáticas.

Em seguida, temos o **Hitchhiker's Guide to the Galaxy**, que se destaca pela capacidade de analisar frases escritas pelo jogador, demonstrando uma notável habilidade na análise léxica.

Por último, o **CodeCombat**, que, embora tenha uma gramática bem definida para aceitar comandos, não foca tão intensamente nesse aspecto e oferece uma variedade menor em comparação com as duas opções anteriores.

Quanto ao segundo requisito, o potencial lúdico, o **CodeCombat** foi escolhido como o melhor devido à sua ampla variedade de cenários e objetivos que podem ser alcançados. Em seguida, o **Hitchhiker's Guide to the Galaxy**, graças à diversidade de desdobramentos de história possíveis, dependendo da criatividade do jogador. Por último, embora possa proporcionar divertimento, o **ANTLRWorks** não oferece tantas opções lúdicas, e não tem um foco tão pronunciado nessa área, em comparação com as outras duas ferramentas.

3. Conclusão

Neste ensaio, propusemo-nos a realizar uma pesquisa centrada em atividades lúdicas fundamentadas em Expressões Regulares e Gramáticas, visando abordar esses conceitos de maneira lúdica, porém educativa.

Em termos de expressões regulares, os resultados da pesquisa foram muito positivos, tendo sido encontradas diversas opções de ferramentas interessantes, que transformam a aprendizagem num processo agradável, simples e interativo.

A principal dificuldade foi procurar ferramentas com propostas interessantes dentro do contexto das gramáticas. Os resultados iniciais desta pesquisa não se revelaram satisfatórios, exigindo uma abordagem alternativa focada na funcionalidade das ferramentas. Esta nova perspectiva trouxe resultados mais positivos, permitindo-nos avaliá-las de uma maneira mais promissora.

Em última análise, o grupo encontra-se satisfeito com os desfechos alcançados ao longo da pesquisa e com as aprendizagens significativas que foram obtidas durante a mesma.

References

- [1] Regulex Site
- [2] Repositório do projeto Regulex
- [3] RegexOne Site
- [4] Regex Golf Site
- [5] Regex Crossword Site
- [6] Hitchhiker's Guide to the Galaxy, the game 30th anniversary edition
- [7] CodeCombat Site
- [8] Organização do CodeCombat
- [9] ANTLWorks Site