

José Alberto Martins Fernandes

A inteligência artificial na gestão do conhecimento



Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
Fevereiro de 2022

José Alberto Martins Fernandes

A inteligência artificial na gestão do conhecimento

Relatório de Estágio Curricular

Orientador: João Pedro Pedroso Ramos dos Santos

Coorientador: Manuel Leite Ribeiro

Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
Fevereiro de 2022

Agradecimentos

Começo por agradecer ao professor João Pedro Pedroso e à Faculdade de Ciências da Universidade do Porto pela oferta deste primeiro passo na minha vida profissional.

Também sou grato às pessoas que me acompanharam neste percurso académico, família e amigos, pelo apoio incondicional que me forneceram sempre que precisei.

Finalmente, agradeço a todos os membros da Cleva Inetum que me acolheram de braços abertos e me ajudaram neste grande salto da minha vida.

Muito obrigado.

Conteúdo

1	Introdução	5
2	Estado da Arte	6
2.1	Diana	6
2.2	Tecnologias Iniciais	7
2.3	Comunicação	8
3	Descrição do Trabalho Realizado	9
3.1	Introdução à Empresa	9
3.2	Metodologia de Trabalho	9
3.3	Introdução ao Problema	9
3.4	Desenvolvimento do Projeto	10
3.4.1	Inicialização Gráfica	10
3.4.2	Pesquisa e Sistema de Recomendação	10
3.4.3	Apresentação dos Dados	12
3.4.4	Constelação	14
3.4.5	Histórico de Pesquisa	16
3.4.6	Submissão de Novas Perguntas	17
3.4.7	Personalização da Pesquisa	17
3.4.8	Nome do Projeto	18
4	Conclusão	19
5	Apêndices	20
5.1	Script em Python	20
5.2	Código do Projeto	21
	Webgrafia	21

Capítulo 1

Introdução

O mundo empresarial defronta-se, atualmente, com o desafio de informar eficientemente o cliente nos negócios relacionados com o negócio dos Seguros. Um dos aspetos que agrava este obstáculo é o facto de os colaboradores estarem, por norma, mais focados na sua função do que no negócio propriamente dito.

Em alguns casos, a dimensão da curva de aprendizagem no negócio também é bastante acentuada, induzindo os colaboradores à dependência de outros colegas para a transmissão e confirmação de conhecimentos. Para minimizar este efeito, torna-se necessária uma solução que acelere a aquisição de conhecimento e que permita uma gestão eficaz e eficiente do mesmo.

Para combater este contratempo, a Cleva Inetum decidiu então lançar este desafio à comunidade académica. Empresa conceituada e líder de mercado na área de *software* para seguros, a Cleva Inetum conta com uma estratégia de crescimento e internacionalização que cria grandes oportunidades para os elementos da sua equipa. Presente no mercado há mais de 30 anos, junta-se, em 2019, ao Grupo Inetum passando a ser denominada como Cleva Inetum.

Com este projeto, pretende-se, então, o desenvolvimento de uma solução informática que apoie os colaboradores de uma qualquer empresa na resolução do problema anteriormente retratado.

O presente trabalho demonstrou ser desafiador e muito interessante, tendo oferecido a oportunidade de participar no desenvolvimento de um trabalho bastante importante e aprofundar possíveis soluções, indicando o melhor caminho a seguir.

O percurso escolhido foi a produção de um motor de busca com um sistema de recomendações incorporado na pesquisa. Esta decisão baseou-se no facto de estes tipos de *software* serem os mais usuais na pesquisa de qualquer tipo de informação. O sistema de recomendações viria a ser um dos fatores essenciais e destacados neste projeto, pois permitiria não só apresentar dados investigados pela empresa, mas também fazer ligações entre os diversos conceitos, possibilitando a uma procura mais eficiência de vários conceitos dentro do mesmo tema.

Capítulo 2

Estado da Arte

O programa que foi desenvolvido ao longo deste trabalho utiliza conceitos de ferramentas já implementadas na Cleva Inetum, tais como a apresentação de documentação local consoante o conceito questionado.

No entanto, não havia limitação na demonstração dessas informações, sendo então essencial repensar a metodologia da ferramenta para endereçar este ponto fulcral.

Neste capítulo é apresentada em detalhe esta ferramenta que se pretende substituir: a Diana.

2.1 Diana

Tal como afirmado anteriormente, a Cleva Inetum tinha acesso a uma ferramenta que, embora melhorasse a experiência do cliente, não o fazia da forma mais eficaz. Essa ferramenta denomina-se Diana.



Figura 2.1: *Diana*

A Diana é um *chatbot* bastante simples que se baseia na deteção de uma palavra conceito que se encontre na base de dados fornecida. Caso não detete, pede então para reformular a questão dando uma segunda oportunidade ao *chatbot* de pesquisar possíveis novas palavras. Caso consiga atingir o objetivo, fornece toda uma documentação da palavra detetada, apresentando no final também um conjunto de outras perguntas dentro do tema e conceitos relacionados. Caso não tenha sucesso nas duas buscas anteriores, a Diana apresenta então

um conjunto de temas gerais para escolher e restringir cada vez mais o conceito desejado.

Todo este processo é feito através de ações razoavelmente demoradas e é visualmente pouco atrativo.

Durante este estágio, foi-me ainda informado que a Cleva Inetum no Porto não tinha acesso ao código desta mesma ferramenta, realçando, por isso, outro ponto positivo no projeto para o qual fui escolhido: a capacidade de administração de uma nova ferramenta interna da empresa.

2.2 Tecnologias Iniciais

O ambiente de desenvolvimento foi um editor de texto simples: o *Visual Studio Code*. Esta decisão resumiu-se ao facto de que um simples editor seria mais do que suficiente para tratar da complexidade do projeto, não havendo necessidade de usar um IDE (cuja aplicação seria mais viável num projeto com maior quantidade de ficheiros relacionados).

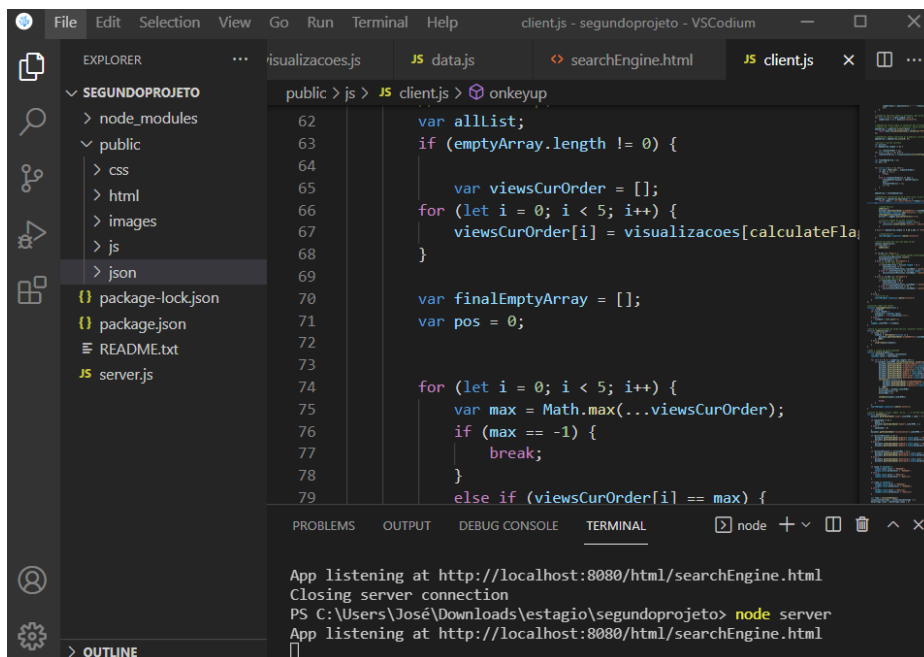


Figura 2.2: *Visual Studio Code*

Numa fase inicial do projeto, decidi que seria mais adequado o uso de *HTML* [1], *CSS* [2] e *Javascript* [3] sem qualquer tipo de adição de bibliotecas. No entanto, depois de uma semana de desenvolvimento de código, deparei-me com problemas de *import* de dados das folhas *Excel* fornecidas.

Devido a essa situação, decidi recomeçar o trabalho adicionando uma nova biblioteca ao

Javascript [3]: o *Node.js* [4]. Esta adição possibilitou o uso de novas funções pré-definidas para transferência e conversão imediata de valores para formato *JSON*, sendo um formato que facilita o uso desses dados necessários.

O *Node.js*[4] não só possibilitou o transporte de dados como também permitiu criar uma estrutura de um servidor de forma mais simples usando o *Express.js* [5].

2.3 Comunicação

Uma das partes mais importantes durante este estágio sempre foi a comunicação. Devido à situação pandémica em que ainda nos encontramos, o estágio teve de ser efetuado maioritariamente em regime remoto. Graças a isso, introduziu-se o uso de videochamada através da aplicação *Microsoft Teams*. Essa adaptação facilitou o contacto constante, sendo benéfico tanto para a empresa, que procurava manter-se atualizada, como para mim, que procurava obter respostas finais quanto a caminhos a seguir no projeto.

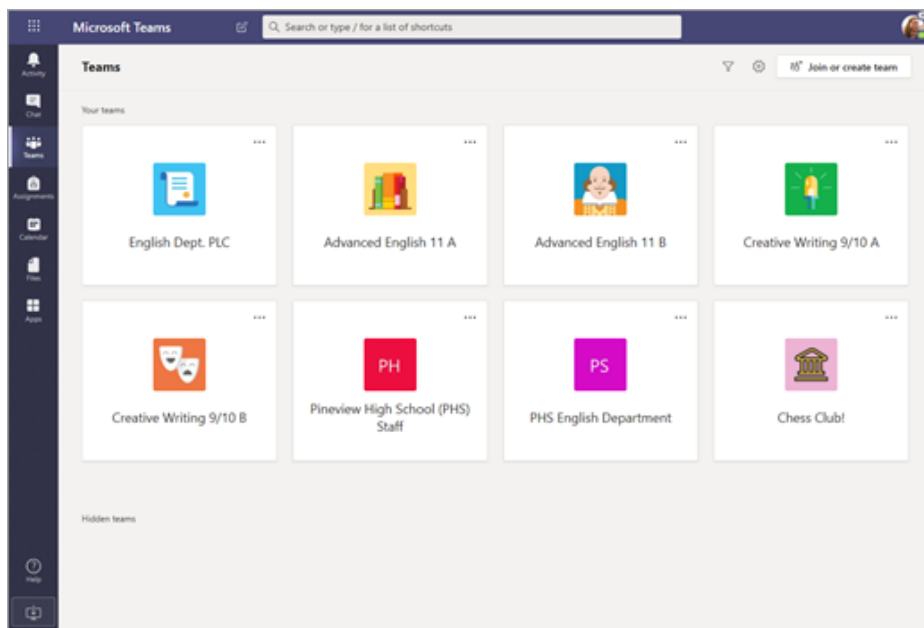


Figura 2.3: *Microsoft Teams*

Capítulo 3

Descrição do Trabalho Realizado

3.1 Introdução à Empresa

No dia 20 de Outubro tive o prazer de visitar pela primeira vez as instalações da Cleva Inetum no Porto. Fui convocado para uma reunião inicial com algumas das pessoas que me iriam acompanhar durante o decorrer deste projeto. Tive a oportunidade de conhecer melhor os meus colegas e as condições disponibilizadas pela Cleva Inetum.

Foi-me fornecido um *kit* para novos membros, um *laptop* para trabalhar neste projeto e, posteriormente, tratou-se dos documentos em falta para oficializar o contrato de estágio curricular.

Fui também adicionado às redes de comunicação da empresa, visto que grande parte deste projeto seria feito de forma remota.

3.2 Metodologia de Trabalho

No dia 25 de outubro, participei no primeiro "Ponto de Situação". Estes encontros, marcados para todas as segundas-feiras até à conclusão do estágio, serviriam para a empresa acompanhar regularmente o meu trabalho durante o desenvolvimento do projeto e para promover uma comunicação aberta.

3.3 Introdução ao Problema

Foi-me então apresentado no dia 25 de outubro o problema que a empresa queria resolver: um novo projeto feito de raíz que serviria de posto informativo sobre o tema *Seguros*, tanto para funcionários da Cleva Inetum, como para os seus diversos clientes. Pediram-me então para investigar, procurar uma solução e apresentar o mais rapidamente possível.

Depois de uma pesquisa das várias possibilidades, apresentei à equipa no dia 8 de novembro um protótipo muito básico do conceito que queria desenvolver: um *search engine* com um sistema de recomendação de temas.

Depois de tirar todas as dúvidas sobre o propósito, o coorientador aprovou a proposta e, conjuntamente com o resto da equipa, apresentou diversas ideias sobre como proceder.

3.4 Desenvolvimento do Projeto

3.4.1 Inicialização Gráfica

Comecei então no dia 9 de novembro a trabalhar no projeto em si. Criei um *frontend* que, embora fosse rudimentar, seria capaz de permitir visualizar o progresso do *backend* de uma forma gráfica. Para a produção da parte estilística, baseei-me em diversos projetos disponíveis no *GitHub*, com o intuito de conhecer mais sobre essa parte da programação da qual não tinha tanto conhecimento, e no *source code* de motores de busca já existentes (não pelo conteúdo em si, mas sim pela estruturação). As fontes usadas foram as fornecidas pela API da *Google* e os ícones usados foram fornecidos pela *Font Awesome*.



Figura 3.1: *Homepage do projeto*

3.4.2 Pesquisa e Sistema de Recomendação

Após ter uma parte gráfica funcional, inicializei uma das componentes fundamentais do projeto: o sistema de procura. Comecei por implementar código que permitisse recomendar temas consoante o que fosse escrito na barra de texto. Essa recomendação inicialmente seria feita por ordem alfabética, sendo que a busca personalizada viria mais tarde.

A recomendação era feita com base em dois ficheiros: o das perguntas e o dos dados (mais especificamente os atributos "conceito"). Assumia-se então que a dúvida do cliente tanto podia ser a pesquisa mais comum (procura simples sobre o conceito) como uma pesquisa mais detalhada (procura de dois fatores usando a junção de uma pergunta com o conceito em questão).

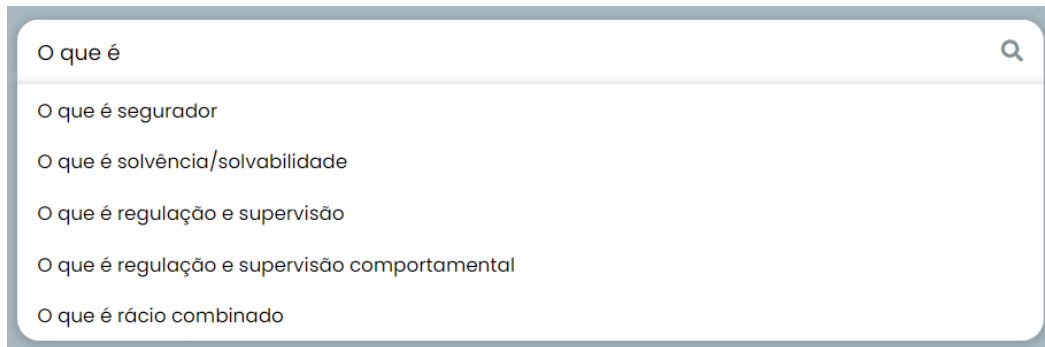


Figura 3.2: Pesquisa do tema com pergunta associada

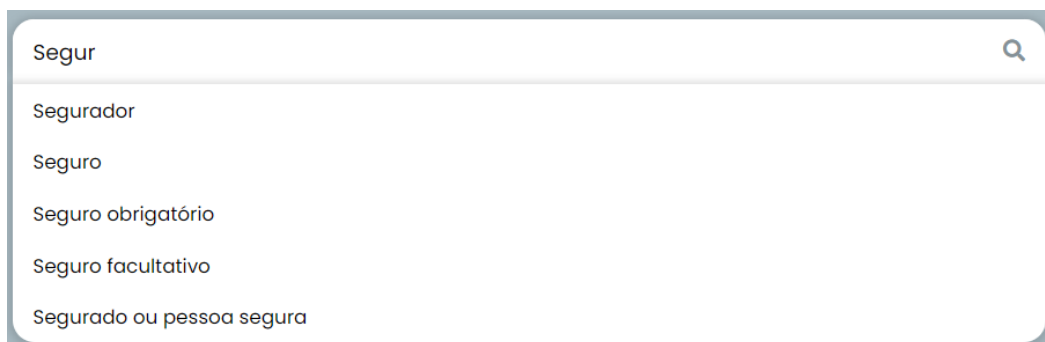


Figura 3.3: Pesquisa do tema sem pergunta associada

O ficheiro das perguntas é uma lista com implementação manual (até à data) em que se procura inserir todas as formas de como um dos utilizadores poderia inserir a sua dúvida, como por exemplo "O que significa?", "O que é?" e "Para que serve?".

O ficheiro dos dados é uma outra lista de implementação manual mas preenchida com o auxílio de um *script* em *Python* e de um tratamento de dados automatizado no carregamento do servidor. Cada elemento nessa lista é um objeto composto pelos atributos que representam o nome do conceito, a sua descrição, a sua *string* de identificação, os termos relacionados e o respetivo tipo de ligação ao conceito.

Este *script* foi criado com o objetivo de retirar todo o conteúdo essencial da única base de dados existente para este projeto. A base de dados é então um *workbook* em formato *Excel* que contém colunas de dados, tais como "Conceito", "Definição" e "Termos Relacionados". Essas colunas são retiradas através do *script* e redirecionadas para o ficheiro *data.js* para

mais tarde poderem ser usadas de uma forma mais "direta", ou seja, sem uso recorrente do *Excel*. Um dos outros motivos para esta implementação deriva do uso habitual do *Excel* pelos membros da equipa, sendo que a conexão direta a esse ficheiro permitiria atualizações mais simplificadas.

	CONCEITO	DESCRIÇÃO	ID CONCEITO	TERMOS RELACIONADOS
1				
2	Segurador	O segurador (ou seguradora) é uma entidade cuja atividade - regulada por lei e supervisionada pela ASF - consiste em assumir um determinado risco do seu cliente (tomador, pessoa segura, beneficiário...) mediante o pagamento de uma remuneração designada de prémio. É comum as seguradoras colaborarem com empresas parceiras (mediadores, corretores, bancos) e delegarem nestas a venda de seguros aos clientes.	ALL.00001	seguro
3				segurado ou pessoa segura
4				tomador do seguro
5				indemnização
6				risco
7	Solvência/solvabilidade e	nomeadamente erros de gestão, um cálculo insuficiente das provisões ou reservas, ou mesmo subtarifação. Para garantir a sua solvência, as entidades seguradoras devem garantir a existência de um capital mínimo que lhes permita fazer face aos seus compromissos. Para isso, os valores recebidos (prémios) não ficam imobilizados, sendo investidos com vista à sua rentabilização. Só assim se geram rendimentos que, em conjunto	ALL.00002	rácio combinado
8				capital mínimo
9				provisões técnicas
10				contabilidade

Figura 3.4: *Workbook/Base de Dados*

3.4.3 Apresentação dos Dados

O próximo passo lógico após a implementação da *search bar* seria, então, a forma como os dados pedidos seriam apresentados. Depois de várias reuniões em que se discutiu este assunto, procurei implementar o melhor possível algo que se enquadrasse dentro do solicitado. O pedido foi a divisão do resultado em duas partes: o contexto textual (que inclui o título, a descrição e outros dados essenciais) e a "constelação".

O contexto textual seria, então, uma breve demonstração do conceito pesquisado, apresentando então o título, a descrição e outros dados essenciais. Estas informações, retiradas do

ficheiro *Excel* fornecido pela Cleve Inetum, são preenchidas pelo Alexandre Gonzaga e pela Rita Pacheco, que me auxiliaram durante a realização deste estágio.

A "constelação" seria um grafo não-direcionado que apresentaria os termos relacionados do conceito procurado com demonstração de níveis de profundidade variados. Através desse complemento gráfico, seria possível também pesquisar esses conceitos relacionados clicando em cima do nó que contém o termo de interesse. A implementação deste conceito permite não só auxiliar qualquer utilizador na busca de conhecimento do que pesquisou, mas também o insere num ambiente em que visualiza a área de estudo em redor, possibilitando uma maior eficiência de procura dentro da área de interesse.



Figura 3.5: Apresentação dos dados associados a "Segurador"

3.4.4 Constelação

Para implementar graficamente este grafo, decidi investir na biblioteca *D3.js*. Esta biblioteca consiste num conjunto de funções que facilitam a demonstração de dados de uma forma visualmente apelativa com o uso de *HTML*, *CSS*, *SVG* e *Canvas*. Um dos principais fatores que me incentivaram a utilizar esta biblioteca foi o elevado número de utilizadores desta ferramenta e a existência de documentação e outros materiais de auxílio para a rápida aprendizagem da mesma.

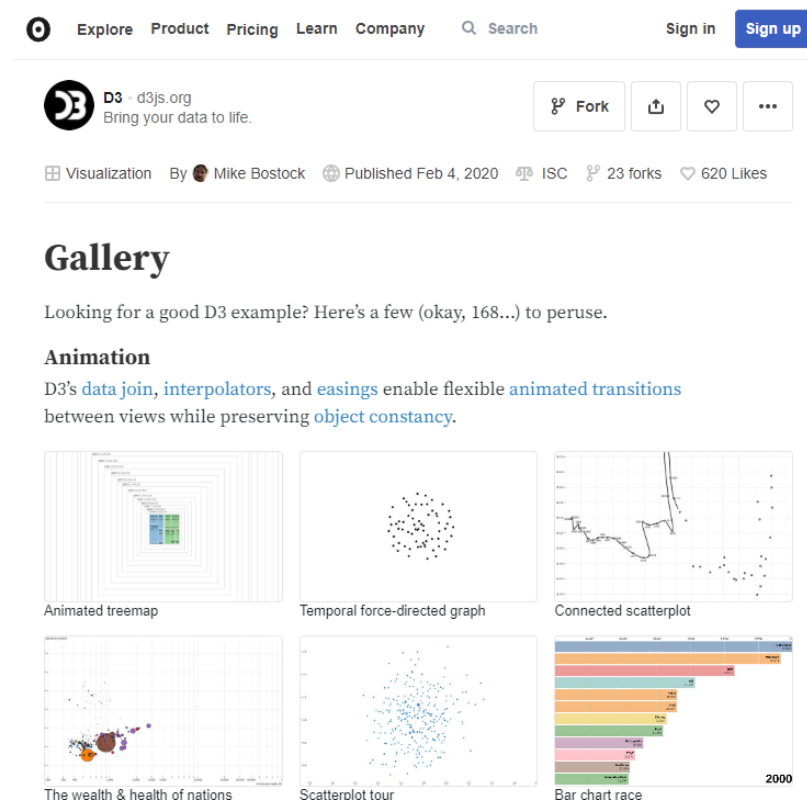


Figura 3.6: *Projetos da comunidade da ferramenta D3.js*

Então, para fazer a representação gráfica da pesquisa, comecei por definir possíveis visualizações. Decidi fazer apenas três níveis de profundidade visto que esse último nível já necessita de mais processamento, demorando mais tempo a ser demonstrado. Graças a isso, qualquer profundidade superior a essa iria retirar um dos elementos principais que este trabalho tenta refletir, a eficiência da pesquisa. Caso o tempo de processamento não fosse de tamanho considerável, não faria sentido na mesma adicionar mais do que três níveis de profundidade, pois a elevada quantidade de nós e ligações diminuiria a facilidade de percepção da área de estudo e, possivelmente, envolveria temas desnecessários com relativamente pouca conexão ao tema principal.

Comecei, então, por processar os conceitos correspondentes ao nível de profundidade desejado, fazer uma lista das ligações. Depois, defini as configurações do grafo. Dentro das configurações do grafo, pude alterar definições relativas à caixa que o contém, aos nós, às ligações, ao posicionamento dos mesmos e para tornar esta visualização mais interativa, às forças aplicadas aos elementos. Também adicionei a possibilidade de clique nos nós para iniciar uma nova pesquisa, a apresentação de texto que demonstre o conceito associado, e animações em caso de *hover* dos nós.

Para navegar entre os três níveis de profundidade, adicionei dois botões e uma caixa de texto com a profundidade atual. Com o clique dos botões, é possível, assim, subir ou descer na profundidade de termos relacionados.



Figura 3.7: *Botões de profundidade*

Um dos problemas que fui encontrando depois de vários testes é a desorganização dos elementos quando em elevado número. Para facilitar a visualização, resolvi implementar um sistema de cores, em que cada nó e as ligações do qual esse nó é "pai" são todas da mesma cor. Essa atualização, facilitou a percepção dos caminhos e tornou a pesquisa mais clara.

Outro problema é o posicionamento dos elementos. Para possibilitar a visualização clara do grafo, foi criada uma caixa consideravelmente grande mas limitada à página *web*. A navegação é, então, feita através das *scrollbars*. No entanto, como o grafo inicia sempre no centro da caixa e as *scrollbars* por *default* começam na posição zero, os elementos não seriam imediatamente visíveis. Para resolver esta situação, a cada alteração e inicialização da caixa inicio as *scrollbars* numa posição centrada.

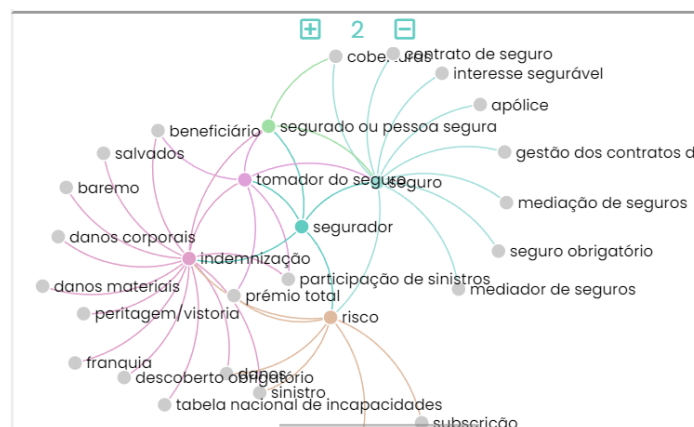


Figura 3.8: *Demonstração do 2º nível de profundidade*

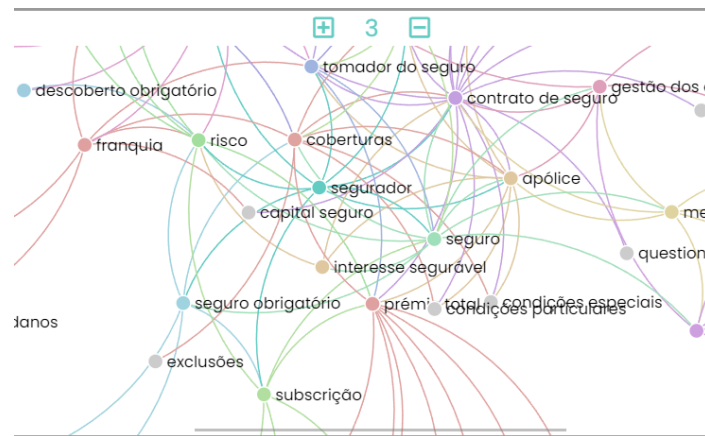


Figura 3.9: Demonstração do 3º nível de profundidade

3.4.5 Histórico de Pesquisa

Um dos principais objetivos que induziu a criação deste projeto era tentar conseguir uma pesquisa de temas mais eficiente para o utilizador. Para cumprir este requisito, a *webpage* não só tinha de fornecer os dados pedidos de uma forma sucinta, mas também dar a possibilidade ao utilizador de pesquisar os vários temas associados à pesquisa efetuada. Este aspeto gerou a necessidade de criar pesquisas enraizadas no tema procurado, mas que pudessem expandir-se para diversos ramos, permitindo ainda retornar a pesquisas anteriores e posteriores.

O utilizador poderia, então, pesquisar termos relacionados através do grafo dirigido. Com o clique num dos nós de interesse, a pagina reencaminha o cliente para a página da nova pesquisa, mantendo um histórico do ramo de pesquisas que está a ser traçado. Foi, então, necessário criar uma estrutura que permitisse a navegação dentro desse histórico. Para isso, foram criados três botões com o intuito de solucionar este problema: os botões "Fechar", "Recuar" e "Avançar".



Figura 3.10: Botões de controlo do histórico

O botão "Fechar" serviria para fechar a procura atualmente aberta e para efetuar um *reset* no histórico, o botão "Recuar" para regressar à pesquisa efetuada anteriormente e o botão "Avançar" para, após recuar no ramo de pesquisas, poder voltar às pesquisas posteriores.

3.4.6 Submissão de Novas Perguntas

Como este projeto se baseia numa base de dados que, embora seja consideravelmente grande, não contempla todos os temas possíveis dentro da área *Seguros*, um ponto importante deste trabalho seria manter a continuidade do desenvolvimento dessa fonte de informação. Necessitou-se, então, de um método que pudesse informar a Cleva Inetum de temas pesquisados, mas aos quais o sistema ainda não conseguia responder.

A solução encontrada foi a criação de um botão de submissão onde, depois de digitar a dúvida para a qual não obteve resposta, o utilizador pudesse clicar para submeter a sua questão para a Cleva Inetum. Essa dúvida seria encaminhada para um ficheiro *.txt* acessível ao *host* do servidor para futura análise e estudo e para mais tarde poder ser inserida na base de dados.



Figura 3.11: *Conceito desconhecido*

3.4.7 Personalização da Pesquisa

No final contratual do estágio, implementei uma parte importante do trabalho: a personalização global da pesquisa. Baseia-se na recolha de informações de pesquisa para melhorar buscas futuras. Devido ao tempo limitado para a realização do projeto, a única personalização implementada foi a contagem de visualizações de cada tema.

Portanto, sempre que uma pesquisa é efetuada, seja na barra de pesquisa ou no clique de um nó do grafo, é atualizado o valor total de pesquisas efetuadas desse conceito. A atualização é feita no ficheiro *data.js* e no *visualizacoes.js*. A necessidade da atualizações dos dois ficheiros

provém do facto de o `data.js` ser atualizado sempre que se reinicia o servidor, precisando de uma base de dados fixa tal como o excel, sendo essa base de dados o ficheiro `visualizacoes.js`.

Com esta personalização, a ordem dos conceitos apresentados após a filtração consoante o que utilizador escreve é feita primeiramente pela contagem de visualizações e depois por ordem alfabética.

Outras personalizações possíveis poderiam ser, por exemplo, a apresentação de temas que não correspondam diretamente com o que o utilizador escreve mas sim com pesquisas efetuadas posteriormente à mais visualizada ou apresentações de temas que não correspondam novamente ao que o utilizador escreve mas a termos relacionados ao conceito mais visualizado.

3.4.8 Nome do Projeto

Um dos desafios finais introduzidos pela Cleva Inetum foi a atribuição de um nome a este motor de busca. Este nome deveria representar a ferramenta e as suas qualidades.

Decidi então chamar este projeto "Atena", uma referência à deusa da sabedoria, procurando representar o objetivo global deste *website*: acesso rápido e eficiente à informação.

Capítulo 4

Conclusão

Este relatório descreve o desenvolvimento de uma *webpage* que procura tornar a pesquisa de informação dentro da área dos seguros mais eficiente e intuitiva.

Este projeto permitiu-me continuar a desenvolver os meus conhecimentos de linguagens que já tinha utilizado previamente tais como *JS* [3], *CSS* [2] e *HTML* [1], e também introduziu linguagens e ferramentas com as quais nunca tinha trabalhado tais como *Python* [6], *Node.js* [4], *D3.js* [7] e *Font Awesome* [8]. Foi também uma oportunidade para desenvolver as minhas capacidades de comunicação com uma equipa brilhante como a que me acompanhou e participar num projeto interessante de uma empresa de renome.

Porém, o aspeto mais importante que pude retirar deste estágio foi a transição dos conhecimentos que adquiri no ensino superior para um contexto profissional. Permitiu-me visualizar uma das possíveis áreas em que, caso tudo corra bem, no futuro poderei vir a trabalhar.

Capítulo 5

Apêndices

5.1 Script em Python

```
import pandas as pd

if __name__ == "__main__":
    #import excel file
    xlsx = pd.ExcelFile('excel.xlsx')

    #divide the existing sheets
    df1 = pd.read_excel(xlsx, 'TERMOS_GERAIS')
    df2 = pd.read_excel(xlsx, 'RAMO_VIDA')
    df3 = pd.read_excel(xlsx, 'RAMO_NAO_VIDA')
    df4 = pd.read_excel(xlsx, 'COSSEGURO_E_RESSEGURO')

    df4 = df4.rename(columns = {"CONCEITO_(PORTUGUES)": "CONCEITO"})

    #export data in json files
    with open("termosGerais.json", "w", encoding='utf-8') as outfile:
        df1.to_json(outfile, force_ascii=False)
        outfile.write('\n')
    with open("ramoVida.json", "w", encoding='utf-8') as outfile:
        df2.to_json(outfile, force_ascii=False)
        outfile.write('\n')
    with open("ramoNaoVida.json", "w", encoding='utf-8') as outfile:
        df3.to_json(outfile, force_ascii=False)
        outfile.write('\n')
    with open("resseguro.json", "w", encoding='utf-8') as outfile:
        df4.to_json(outfile, force_ascii=False)
        outfile.write('\n')
```

5.2 Código do Projeto

Devido à quantidade elevada de código desenvolvido durante este estágio, disponibilizo o mesmo no seguinte *link*:

<https://github.com/josefernandes99/Atena-SearchEngine>

Webgrafia

- [1] “Hypertext markup language (html).” <https://html.spec.whatwg.org/multipage/>.
- [2] “Cascading style sheets (css).” <https://www.w3.org/Style/CSS/Overview.en.html>.
- [3] “Javascript (js).” <https://www.javascript.com/>.
- [4] “Node.js.” <https://nodejs.org/en/>.
- [5] “Express.js.” <https://expressjs.com/>.
- [6] “Python (py).” <https://www.python.org/>.
- [7] “D3.js.” <https://d3js.org/>.
- [8] “Font awesome.” <https://fontawesome.com/>.