



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA INFORMÁTICA

CURSO DE ENGENHARIA INFORMÁTICA

Disciplina: *SISTEMAS DISTRIBUÍDOS*

Googol

PROJETO SISTEMAS DISTRIBUÍDOS

GOOGOL

Regente da disciplina

Professor Raul André Barbosa

Trabalho realizado pelos alunos:

Leonardo Moreira Pina (2019234318)

Luís Filipe Neto (2020215474)

Pedro Guilherme Silva Santos das Neves Ascensão (2020233012)

- **Índice**

	Pág.
• Introdução	1
• Arquitetura do sistema	1
• Interface Web	1
• /login	2
• /	2
• /pesquisar	3
• /info	3
• /indexar	4
• /indexarStories	4
• /consultar	5
• Serviço REST	5
• Testes realizados	6
• Conclusão	7

- **Introdução**

Este projeto foi realizado pelos alunos Leonardo Pina (nº 2019234318), Pedro Ascensão (nº 2020233012) e Luís Neto (nº 2020215474) com o objetivo de criar um motor de pesquisa de páginas Web, sendo que nesta segunda meta utilizámos as funcionalidades desenvolvidas na meta anterior e desenvolvemos uma interface Web, de forma a complementar estas funcionalidades, utilizando as diferentes ferramentas propostas pelo professor.

- **Arquitetura do sistema**

Desta forma, foi seguida a arquitetura apresentada no enunciado, arquitetura encontra-se bastante assente na presente na meta 1. Vejamos então agora o esquema da arquitetura:

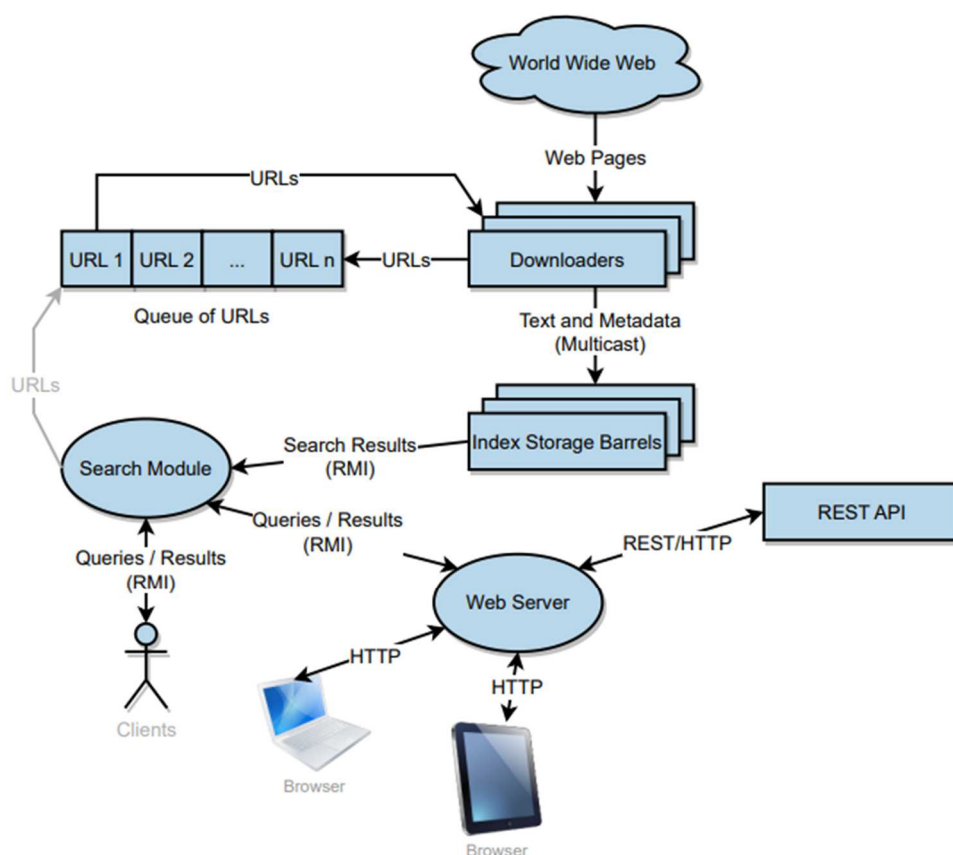


Fig. 1 – Arquitetura do sistema

- **Interface Web**

Existem, portanto, várias vistas disponíveis que apresentam as funcionalidades do motor de busca desenvolvidas na meta anterior.

Assim, a classe **Controlador** está encarregue de controlar todas as vistas e estabelecer todas as ligações entre estas (os ficheiros html) e as classes provenientes da meta anterior, assim como as ligações RMI existentes entre estas duas componentes.

Após o utilizador ter todas as aplicações devidas a correr corretamente, ao aceder ao localhost:8080 está pronto para explorar o website. Iremos agora então ver mais detalhadamente o que cada uma destas vistas está encarregue de mostrar:

- **/login**

Ao ingressar no website, o utilizador é encaminhado para esta vista, no qual é necessário fazer login com uma conta já existente, ou caso o utilizador ainda não possua nenhum, deve então proceder à criação de uma nova conta, efetuando o registo.

Desta forma, é feita uma ligação RMI com o search module de forma a verificar os campos de login ou registo completados.

Googol

Seja bem-vindo ao seu motor de busca favorito!!

The image shows two side-by-side forms. The left form is titled 'Login' and contains fields for 'Nome de utilizador' and 'Password', with a blue 'Login' button at the bottom. The right form is titled 'Registo' and contains fields for 'Nome', 'Nome de utilizador', 'Email', and 'Password', with a blue 'Registrar' button at the bottom.

Fig. 2 – Página de login e registo

- **/**

Após o login ou registo estar concluído, o utilizador é encaminhado para a página “inicial”, onde pode optar por escolher várias opções de páginas, consoante a funcionalidade que desejar realizar. Consoante a opção tomada, este irá ser levado para outra vista dedicada à funcionalidade que o mesmo estará prestes a explorar.

The image shows the initial page of the Googol website. It features the 'Googol' logo and the text 'Seja bem-vindo ao seu motor de busca favorito!!'. Below this, there are five buttons, each with a blue 'Ir' button: 'Indexar URL', 'Indexar Hacker News Stories', 'Pesquisar', 'Consultar lista de páginas', and 'Obter informações gerais'.

Fig. 3 – página inicial

- **/pesquisar**

Ao escolher esta página, o utilizador poderá então pesquisar por uma palavra, usando a search bar existente, sendo que serão de seguida devolvidas todos os urls (links) de todas as páginas web que contêm essa palavra.

Assim, existe, no entanto, uma check box encarregue de filtrar esta pesquisa, na medida que se esta for selecionada ao realizar a pesquisa da palavra, a aplicação irá mostrar não só os urls já indexados, ou seja, existentes nos storage barrels, mas também as top stories dos Hackernews que satisfaçam a dada pesquisa.

Na classe **Controlador** já referida, o bloco correspondente a gerar esta vista comunica por uma ligação RMI com o search module, de modo a poder receber estas mensagens vindas dos storage barrels, onde pesquisa por estas páginas. Ao mesmo tempo, é importante referir que as top stories encontradas no Hackernews (caso o botão tenha sido selecionado) serão também de seguida indexadas para a fila url onde serão por sua vez inseridas nos devidos storage barrels.

#	ID	Título	URL	Autor
1	35978151	Retool (YC W17) is hiring hackers to combine visual programming with AI	---	jacqt
2	35974037	Show HN: Subscriber-Only – Paid Subscriptions for Jekyll	https://subscriber-only.com	vmisp
3	35973645	OpenAI came after our domain because we use GPT in it	---	decide1000
4	35976415	Tech stack for fine-tuning LLMs	---	behat
5	35964590	Show HN: DevPod – Codespaces but Open Source, Client-Only, and Unopinionated	https://github.com/loft-sh/devpod	gentele
6	35964640	Show HN: Neucards – Privacy based digital contact card	https://www.neucards.com	bdominy
7	35948462	Show HN: Willow – Open-source privacy-focused voice assistant hardware	https://github.com/toverainc/willow	kkiehofner
8	35958715	Show HN: Oblivus GPU Cloud – Affordable and scalable GPU servers from \$0.29/hr	https://oblivus.com	oblivuslimited

Fig. 3 – Página Pesquisar

- /info

Esta página tem como objetivo exibir as informações gerais do sistema, em tempo real, tal como ocorreu na primeira meta. Para conseguirmos uma atualização em tempo real, foi necessário utilizar um WebSockets. Assim, quando o utilizador abre a página é criado o WebSockets, através da função connect(), localizada no html desta página. Ademais, este utilizador também fica inscrito num endpoint, local onde irão as mensagens recebidas através do search module.

Para obtermos uma atualização em tempo real, a abordagem utilizada consistiu numa função que se comunica através de RMI com o Search Module, solicitando todas as informações necessárias. Ainda importa acrescentar que esta função é chamada em cada segundo e localiza-se na classe **Controlador**.

Googol Indexar Indexar Stories Pesquisar Consultar Informações Logout			
<h2>Informação Geral</h2> <p>Aqui poderá obter toda a informação sobre a nossa plataforma!</p>			
Storage Barrels			
#	IP	Porto	Gama
1	192.168.1.95	9876	[A-M]
Downloaders			
#	IP	Porto	ID
1	192.168.1.95	4322	Downloader1
Pesquisas frequentes			
#	Pesquisa	Frequência	
1	energia	35	
2	potetu	8	
3	if	7	
4	https://www.uc.pt	7	

Fig. 4 – Página Informação Geral

- /indexar

Esta vista tem como objetivo introduzir um URL para ser indexado, dado pelo utilizador. Deste modo, foi utilizada a ligação RMI já estabelecida anteriormente com o Search Module para chamar a função que trata de introduzir o URL na fila de URL's a serem indexados.

Googol	Indexar	Indexar Stories	Pesquisar	Consultar	Informações	Logout
<h2>Indexar URL</h2> <p>Aqui poderá adicionar um URL que pretenda que o nosso motor de busca processe!</p>						
<input type="text" value="www.uc.pt"/>		<input type="button" value="Indexar"/>				
URL adicionado com sucesso!						

Fig. 5 – Página Indexar

- /indexarStories

Esta página será utilizada para o utilizador poder indexar todas as Stories que um certo utilizador submeteu. Para esta funcionalidade foram então utilizadas as funções de REST já descritas no tópico Serviço REST, sendo depois enviados os URL's, por RMI, para o Search Module.

Googol

Indexar

Indexar Stories

Pesquisar

Consultar

Informações

Logout

Indexar Hacker News Stories

Aqui poderá procurar por um utilizador e obter todas as suas stories!

jl

Indexar

Adicionados 41 de 701 urls com sucesso!

Resultados da pesquisa - Hacker News

#	ID	Autor	Título	URL
1	35675818	jl	Listen to Steve Huffman tell the story of how Reddit got started	https://pod.link/1677066062/episode/e1eed71375798a8850bbfd90b03256bc
2	19464269	jl	Women: Learn to Program This Summer	http://foundersatwork.posthaven.com/women-learn-to-program-this-summer
3	10335836	jl	AMA with one of the pro bono defense attorneys for Richard Glossip	https://www.reddit.com/r/IAmA/comments/3nmz8w/i_am_don_knight_one_of_the_pro_bono_defense/
4	10216820	jl	XQ: The Super School Project	http://www.nytimes.com/2015/09/14/us/laurene-powell-jobs-commits-dollar50-million-to-create-new-high-schools.html?_r=0
5	9097991	jl	FFC: Adora Cheung of Homejoy (YC S10) on Scaling Your	https://www.youtube.com/watch?v=Z9C1k8mEdtY&index=7&list=PLQ-uHSnFig5NECe5MbTzqJaVjxoEjunvB

Fig.6 – Página Indexar Hacker News Stories

- /consultar

A vista consultar tem como propósito listar todas as páginas que tem ligação para a página que o utilizador introduziu no form desta página. Este serviço é mais uma vez realizado através da ligação RMI.



#	URL
1	https://www.uc.pt

Fig. 7 – Página Consultar

- Serviço REST

Na realização deste projeto, utilizamos um serviço REST, integrando-o na aplicação de forma a podermos obter as Hackernews nas pesquisas realizadas.

Desta forma, ao estudarmos a documentação <https://github.com/HackerNews/API> fornecida pelo professor, utilizámos como forma de auxílio o Postman, isto para podermos perceber que estrutura de dados cada link devolvia.

De forma a realizarmos as funcionalidades pedidas, na classe **MyHackerNewsController**, inicializámos o REST cliente, sendo que para a primeira funcionalidade: fornecemos o url que contém todas as top stories existentes; é devolvido um objeto JsonNode que contém uma lista com o id (numérico) de todas estas stories; percorremos esta lista e para cada id obtido fizemos uma nova ligação, inicializando um novo REST cliente, no qual fornecemos o url que contém esse dado item, ou seja, essa dada story; ao serem devolvidas informações acerca dessa story e tal como especificado no enunciado, procurámos no texto a ocorrência da palavra dada pelo utilizador na pesquisa (rejeitando todas as stories que não contivessem essa palavra ou que não tivessem sequer o campo “text”); devolvemos por último todas as stories obtidas na API, de forma a poderem ser posteriormente usadas e mostradas nas devidas vistas apresentadas pela aplicação. Para a segunda funcionalidade (que corresponde ao endpoint **/indexarStories**) o processo foi bastante semelhante: fornecemos o url que contém o id do utilizador cujas stories submetidas queremos encontrar, id este fornecido pelo utilizador; é devolvido um objeto JsonNode que contém várias informações acerca do user pesquisado, no qual iremos percorrer a lista presente no campo “submitted”, campo este que apresenta todos os id’s de todas as stories submetidas por este user; para cada id encontrado, fizemos uma nova ligação, inicializando um novo REST cliente, no qual fornecemos o url que contém esse dado item; ao serem devolvidas informações acerca desse item, tal como especificado no enunciado, verificámos se se tratava de uma story (uma vez que só queremos as stories submetidas e não outro tipo de notícias); por último devolvemos todas as stories obtidas na API obtidas nesta filtragem, de forma a poderem ser posteriormente usadas e mostradas nas devidas vistas apresentadas pela aplicação.

• Testes realizados

Após ter sido desenvolvida a arquitetura do sistema, foram definidos então alguns testes de forma a ser possível testar o bom funcionamento do programa como um todo e ainda, se o seu principal objetivo foi cumprido, vejamos agora uma tabela com alguns desses testes realizados:

ID	Teste realizado	✓ / X
1	Fazer login com uma conta não registada resulta em erro	✓
2	Fazer login com uma conta registada dá acesso à aplicação	✓
3	Fazer registo com credenciais novas dá acesso à aplicação	✓
4	Fazer registo com credenciais repetidas resulta em erro	✓
5	Na página principal existem métodos de acesso a todas as funcionalidades disponibilizadas	✓
6	Cada um dos métodos de acesso funcionam e redirecionam para páginas diferentes	✓
7	Na página “Indexar URL”, ao premir o botão “Indexar” a função de indexar apenas é chamada se o campo estiver preenchido	✓
8	Na página “Indexar URL”, ao premir o botão “Indexar” a aplicação indexa o URL devidamente na Fila	✓
9	Na página “Indexar URL”, ao premir o botão “Indexar” a aplicação apresenta um erro caso não tenha sido possível indexar o URL devidamente na Fila	✓
10	Na página “Indexar Stories”, ao premir o botão “Indexar” a função de indexar apenas é chamada se o campo estiver preenchido	✓
11	Na página “Indexar Stories”, ao premir o botão “Indexar” a aplicação indexa todos os URL de todas as stories do utilizador encontrado na Fila	✓
12	Na página “Indexar Stories”, ao premir o botão “Indexar” a aplicação apresenta um erro caso o utilizador não tenha sido encontrado	✓
13	Na página “Pesquisar”, ao premir o botão “Procurar” a aplicação apresenta todos os resultados da pesquisa inserida	✓

14	Na página “Pesquisar”, ao premir o botão “Procurar” a aplicação apresenta uma mensagem de erro caso não tenham sido encontrados resultados para a pesquisa inserida	✓
15	Na página “Pesquisar”, caso a checkbox “Ativar Hacker News” esteja ativada, ao premir o botão “Procurar” a aplicação apresenta uma mensagem de erro caso não tenham sido encontrados resultados para a pesquisa inserida	✓
16	Na página “Pesquisar”, caso a checkbox “Ativar Hacker News” esteja ativada, ao premir o botão “Procurar” a aplicação apresenta todos os resultados da pesquisa inserida juntamente com os resultados vindos da Hacker News	✓
17	Na página “Consultar”, ao premir o botão “Procurar” a aplicação apresenta todos os resultados da pesquisa inserida	✓
18	Na página “Consultar”, ao premir o botão “Procurar” a aplicação apresenta uma mensagem de erro caso o URL não tenha sido encontrado	✓
19	Na página “Informação Geral”, a aplicação apresenta todas as informações referentes aos Storage Barrels, Downloaders e mapa de pesquisas em tempo real	✓
20	Na página “Informação Geral”, caso não exista nenhuma informação sobre Storage Barrels, Downloaders ou mapa de pesquisas apenas aparecerá o header da tabela sem informação por baixo	✓
21	Em todas as páginas existe a possibilidade de fazer Logout com o botão existente na barra de navegação	✓
21	Ao fazer Logout, o utilizador deixa de ter acesso à aplicação, podendo voltar a ter acesso se fizer novamente o Login	✓
22	Na página “Pesquisar”, caso a checkbox “Ativar Hacker News” esteja ativada, ao premir o botão “Procurar” a aplicação indexa todos os URLs de todos os resultados obtidos e mostra uma mensagem com o número de URLs indexados na fila	✓

• Conclusão

Concluimos dizendo que durante a resolução destes projetos deparámo-nos com vários problemas, maioritariamente na integração de web sockets e, portanto, atualizações em tempo real. No entanto, pensamos ter conseguido resolver as tarefas propostas e o projeto proposto, sendo que conseguimos obter com sucesso um resultado final de algo que se esperaria que fosse um motor de pesquisa de páginas web.

Referimos ainda que este projeto nos ajudou melhor a compreender como juntar diferentes componentes de um projeto deste tipo, assim como as ligações necessárias ao seu bom funcionamento e é ainda importante mencionar que a sua resolução nos ajudou a conseguir melhor perceber os conceitos lecionados nas aulas, como os implementar e complementar num contexto prático.