

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Spidious: Improving User's Efficiency on Email with Machine Learning and a Better Visualization Approach

Tiago Filipe Ferreira Costa



FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Informática e Computação

Orientador: Doutor Rui Maranhão

16 de Julho de 2015

Spidious: Improving User's Efficiency on Email with Machine Learning and a Better Visualization Approach

Tiago Filipe Ferreira Costa

Mestrado Integrado em Engenharia Informática e Computação

Aprovado em provas públicas pelo Júri:

Presidente: Doutor José Manuel de Magalhães Cruz

Arguente: Doutor José Maria Amaral Fernandes

Orientador: Doutor Rui Filipe Lima Maranhão de Abreu

16 de Julho de 2015

Resumo

A comunicação está presente na vida humana desde os seus primórdios, inicialmente em pequena escala com recurso a pinturas rupestres e gestos, e mais tarde nomeadamente no século XX, em grande escala através do uso da tecnologia. No entanto, foi sempre um meio fundamental na vida das pessoas que sofreu, e sofre, uma evolução contínua.

Em virtude desta necessidade constante surgiu, em 1969, uma das ferramentas de comunicação mais duradouras e mais usadas: o e-mail. Este, da forma como o conhecemos, está extraordinariamente enraizado um pouco por todo o mundo ao ponto de serem enviados cerca de 196 biliões de e-mails por dia estimando-se ainda que, até 2018, existam aproximadamente 5.2 biliões de contas registadas. Aquando da sua projeção, o e-mail foi pensado para resolver um problema: o de comunicar entre pessoas à distância. E a verdade é que conseguiu resolvê-lo de forma eficaz.

No entanto, à medida que esta solução se tornou mais popular, aliada a um ritmo gritante verificado no trabalho e na vida das pessoas, originou a criação de um outro problema, de maior gravidade, ainda não resolvido: o da falta de eficiência no uso do e-mail. Este mesmo problema potencia a existência de níveis de entropia assustadores nas contas de e-mail e, conseqüentemente, leva a uma perda de tempo importantíssimo para o utilizador, tempo esse que em situações empresariais se materializa em custos elevados.

Note-se que este é um problema com uma importância indescritível e pouco estudado, que indubitavelmente não pode continuar sem uma solução confiável. É estimado que, o e-mail consome cerca de 25% do tempo de trabalho dos seus utilizadores. E se, no passado, o ser humano conseguiu melhorar a sua comunicação presencial dado que a mesma era essencial, sendo a comunicação através do e-mail também ela fulcral no presente, porque não haveremos nós de a melhorar também?

De facto, atualmente, tem-se assistido a um grande interesse por parte das empresas tecnológicas relativamente ao tema do e-mail, interesse esse que, tem levado ao aparecimento de algumas tentativas de solução aos problemas apresentados pelo e-mail. De entre esses podemos destacar, por exemplo, o Inbox da Google, o ActiveInbox ou até mesmo o Boomerang for Gmail da BaydIn. No entanto, e infelizmente para todos os utilizadores, cada uma destas soluções apenas resolve pequenos pormenores ou partes do problema e, por essas razões, nenhuma delas é uma alternativa viável para resolver a questão científica em estudo.

Desta forma, esta dissertação pretende afirmar-se como uma solução assertiva que permita resolver o problema de eficiência atual do e-mail. Para isso, a mesma passará por duas fases distintas. A primeira, e a principal, pela criação de uma nova forma de visualização para o e-mail, assente em novos paradigmas, que dote o utilizador de ferramentas imprescindíveis para gerir facilmente um grande volume de mensagens. E a segunda, de forma complementar com o objetivo de demonstrar a sua aplicabilidade ao problema, a de conferir ao e-mail um determinado nível de inteligência através de modelos estatísticos que se vão moldando com as próprias ações do utilizador. Desse modo pretende-se fazer com que, seja o e-mail a trabalhar para o utilizador e não o utilizador a trabalhar para o e-mail.

Abstract

Communication is present in human life since the beginning, initially at a small scale using the cave paintings and gestures and later, viz. in the twentieth century, at a much larger scale through the use of technology.

In 1969, by the virtue of this constant need, came into existence one of the most durable and more used tools of communication: the email. This, in the way we know it, is deeply rooted all over the world to the point that it is estimated that about 196 billion emails are sent on a daily basis, and it is also estimated that the number of email accounts to be approximately 5.2 billion by 2018. At the time of its projection, the email was thought to solve a problem: to communicate with people at a distance. And the truth is that it was solved effectively.

However, as this solution has become more popular, associated with a very high pace checked at work and in people's lives, it afforded the creation of another problem, more serious, not solved yet: the lack of efficiency in the use of email. This same problem enhances the existence of high levels of entropy in e-mail accounts and, consequently, it leads to a very considerable loss of time for the user, and in business situations materializes in high costs.

Note that this is a problem with an unsurmountable importance and understudied, that certainly cannot continue without a reliable solution. It is estimated that the email consumes about 25% of the working time of its users. And if, in the past, the human being improved its face-to-face communication, once that it was essential, being the communication through email important in the present, why we can't improve that too?

In fact, currently, we have seen a great interest from technology companies, in relation to email subject, that it has led to the appearance of some attempted solutions to the problems posed by email. Among these we can highlight, for example, the Inbox by Google, the ActiveInbox, or even the Boomerang for Google da BaydIn. However, and unfortunately for all users, each one of these solutions solves only small details or parts of the problem and, for these reasons, none of them is an alternative feasible to resolve the scientific question under study.

Thus, this MSc thesis aims to be an assertive solution that solves the problem of current email efficiency. To this, it will go through two distinct phases. The first involves the creation of a new way of viewing for email, based on new paradigms, which will provide the user essential tools to easily manage a large volume messages. And the second, gives the email a certain level of intelligence using statistical models will shaping with their own user actions, causing the email work for the user and not the user for the email, making it a good servant and a good master.

Dedicatória

Dedicado a toda a minha família, aos meus amigos, à minha namorada, mas sobretudo a uma pessoa especial, que se encontra num lugar especial: o meu pai.

(Agostinho Costa, 1953 – 2015)

*”Ohana quer dizer família.
E na família ninguém é deixado para trás ou é esquecido...”*

— Lilo & Stitch

Agradecimentos

A dissertação que se segue resulta da ambição, do trabalho árduo e da colaboração, de diversas formas, de muitas pessoas que de algum modo fazem parte da minha vida. Posso afirmar seguramente que nos tempos que passei na Faculdade de Engenharia da Universidade do Porto tive o privilégio de trabalhar com grandes professores, e com alunos excepcionais que serão certamente dos melhores profissionais da área no futuro.

Em primeiro lugar gostaria de agradecer ao meu orientador, o Professor Rui Maranhão, que me guiou e ajudou, com a sua experiência e conhecimentos, a trilhar este longo caminho. A sua importância é imensurável, uma vez que marcou sem dúvida alguma a minha futura carreira. O meu obrigado professor.

De seguida gostaria de agradecer aos meus amigos de todas as horas, o Filipe Oliveira e o Tiago Azevedo. Não tenho forma de descrever a vossa ajuda, uma vez que na verdade, a forma mais justa que tenho de a caracterizar é dizer que sempre que eu precisei vocês estiveram lá para me apoiar. De certa forma, parte desta dissertação também vos pertence.

Não posso também deixar de agradecer à minha namorada Adriana Rodrigues. Para ti sou parco em palavras, gostaria apenas de te agradecer por todos os momentos em que ficas incondicionalmente do meu lado.

O meu agradecimento também para todos aqueles que participaram no estudo com utilizadores. Assim sendo gostaria de agradecer: à Adriana Rodrigues, ao Alexey Seliverstov, à Ana Sofia Ferreira, ao André Freitas, ao André Rodrigues, ao Professor António Coelho, ao Bruno Lima, ao Carlos Amaral, ao Cristiano Rodrigues, ao Filipe Oliveira, ao Humberto Rodrigues, ao João Reis, ao Joaquim Barros, ao José Magalhães, ao Professor José Magalhães Cruz, à Marta Reis, ao Rodrigo Costa, ao Tiago Azevedo, ao Vasco Gonçalves e Vítor Sousa. É com certeza que digo que vocês fizeram parte de algo verdadeiramente importante.

Por fim gostaria também de agradecer ao meu amigo Bruno Almendra por ter efetuado a revisão literária a esta dissertação. Aliás no capítulo das amizades existem muitas outras pessoas a quem eu obrigatoriamente necessitava de agradecer. Desse modo, deixo também o meu agradecimento a todos os meus amigos e família que, de algum modo, contribuíram para aquilo que eu sou, e consequentemente para esta dissertação também.

Tiago Costa

Conteúdo

1	Introdução	1
1.1	Contexto	1
1.2	Motivação	2
1.3	Questões de Pesquisa	2
1.4	Objetivos e Contribuições Esperadas	3
1.5	Estrutura do Documento	3
2	Fundamentos	5
2.1	Arquitetura do E-mail	5
2.2	Cliente de E-mail	6
3	E-mail	7
3.1	Interação Pessoa-Computador	8
3.1.1	Padrões de Utilização	8
3.1.2	Fluxo de Trabalho	12
3.1.3	Métodos de Visualização	18
3.2	Aprendizagem e Extração de Conhecimento	25
3.2.1	Classificação	25
3.2.2	Sumarização	27
3.3	Trabalhos Industriais	29
3.3.1	SaneBox	30
3.3.2	ActiveInbox	30
3.3.3	Inboxcube	31
3.3.4	Inbox by Gmail	32
3.3.5	Alto by AOL	33
3.3.6	Mailbox	34
3.3.7	Boomerang for Gmail	35
3.3.8	Análise Comparativa	36
4	Inquérito sobre o e-mail aos utilizadores	37
4.1	Metodologia	37
4.2	Resultados	38
4.2.1	Geral	38
4.2.2	E-mail	39
4.2.3	Visualização	42
4.2.4	Inteligência	47
4.2.5	Considerações Finais	50
4.3	Discussão dos Resultados	51

CONTEÚDO

4.4	Resumo	53
5	Spidious: Solução	55
5.1	Problema	55
5.2	Tecnologias	56
5.2.1	HTML5	56
5.2.2	CSS3	56
5.2.3	JavaScript	56
5.2.4	NodeJS	57
5.2.5	EmberJS	57
5.2.6	RethinkDB	57
5.2.7	WebSocket	58
5.3	Arquitetura	58
5.4	Visualização	59
5.5	Inteligência e Aprendizagem	63
5.6	Resumo da Solução Final	64
6	Spidious: Estudo com os utilizadores	69
6.1	Metodologia	69
6.1.1	Cenário	70
6.1.2	Tarefas	71
6.2	Resultados	72
6.2.1	Pré-Questionário	72
6.2.2	Guião de Tarefas	72
6.2.3	Pós-Questionário	74
6.3	Discussão dos Resultados	85
6.4	Resumo	86
7	Conclusões	89
7.1	Resumo	89
7.2	Contribuições	91
7.3	Trabalho Futuro	91
	Referências	93
A	Inquérito sobre o e-mail aos utilizadores	101
B	Spidious: Estudo com os utilizadores	109
B.1	Pré-Questionário	109
B.2	Resultados do Pré-Questionário	116
B.3	Guião de Tarefas	125
B.4	Resultados do Guião de Tarefas	131
B.5	Pós-Questionário	137

Lista de Figuras

2.1	Imagem representativa da arquitetura base do sistema de e-mail.	5
3.1	Representação das duas fases presentes na comunicação via e-mail: da expectativa à quebra. Imagem retirada do artigo de Joshua Tyler e John Tang.	15
3.2	Processo sugerido para lidar com o e-mail segundo Milos Vacek.	17
3.3	Imagem do TimeStore retirada da publicação de Kelvin Yiu et al.	19
3.4	Imagem da interface do Loom retirada da publicação de Judith Donath, Karrie Karahalios e Fernanda Viegas.	19
3.5	Imagem da interface para o e-mail proposta por Steven Rohall et al.	20
3.6	Imagem da interface para o e-mail proposta por Olle Balter e Candace Sidner. . .	21
3.7	Imagem da interface para o e-mail proposta por Jacek Gwizdka.	21
3.8	Imagem da interface para o e-mail proposta por Victoria Bellotti et al.	22
3.9	Imagem da técnica de visualização <i>ThreadArcs</i> proposta por Bernard Kerr.	23
3.10	Imagem da visualização proposta por Gina Venolia e Carman Neustaedter.	23
3.11	Imagem da visualização proposta por Mirko Mandic e Android Kerne.	24
3.12	Imagem da visualização proposta por Victoria Bellotti e Jim Thornton.	24
3.13	Imagem da visualização proposta por Romain Vuillemot, Jean Petit e Mohand Hacid. .	25
3.14	Imagem da interface do SaneBox.	30
3.15	Imagem da interface do ActiveInbox.	31
3.16	Imagem da interface do Inboxcube.	32
3.17	Imagem da interface do Inbox by Gmail.	33
3.18	Imagem da interface do Alto by AOL.	34
3.19	Imagem da interface do Mailbox.	35
3.20	Imagem da interface do Boomerang for Gmail.	36
4.1	Gênero.	39
4.2	Ocupação profissional.	39
4.3	Tipo de cliente de e-mail utilizado.	40
4.4	Clientes de e-mail utilizados.	40
4.5	Local de consulta do e-mail.	41
4.6	Frequência de consulta do e-mail.	41
4.7	Tempo médio diário despendido no e-mail.	41
4.8	Mensagens de e-mail recebidas por dia.	42
4.9	Mensagens de e-mail enviadas por dia.	42
4.10	Número médio de mensagens não lidas na caixa de entrada.	43
4.11	Tipo de mensagens de e-mail deixadas na caixa de entrada.	43
4.12	Número de pastas excluindo as existentes por defeito.	44
4.13	Método de pesquisa de mensagens mais utilizado.	44

LISTA DE FIGURAS

4.14	Critério de pesquisa mais utilizado.	44
4.15	Tipo de notificações automáticas utilizadas.	45
4.16	Classificação das notificações automáticas quanto à sua intrusividade.	45
4.17	Frequência de utilização de uma mensagem já recebida para obter o e-mail do destinatário para uma nova mensagem.	46
4.18	Satisfação dos utilizadores em relação à atual interface dos clientes de e-mail.	46
4.19	Opinião dos utilizadores quanto à prejudicialidade da atual visualização do e-mail face a sua produtividade pessoal.	47
4.20	Nível de satisfação com as funcionalidades de inteligência e aprendizagem já desenvolvidas para o e-mail.	48
4.21	Receção de e-mails sem interesse em ser lidos aquando da sua chegada.	48
4.22	Opinião sobre sobrecarga causada pela quantidade de mensagens de e-mail para processar.	48
4.23	Pretensão por um cliente de e-mail capaz de se adaptar ao perfil do utilizador.	49
4.24	Pretensão por perceber o conteúdo de um conjunto de mensagens antes mesmo de as ler, através da visualização de meta-informação.	49
4.25	Opinião sobre aumento de produtividade e eficiência no uso do e-mail, causado por uma classificação clara das <i>threads</i> em trabalho ou lazer.	50
5.1	Arquitetura da solução proposta.	59
5.2	Listagem de conversas lateral.	60
5.3	Visualização das <i>threads</i> em conversa.	61
5.4	Caixa de resposta rápida a uma <i>thread</i> de e-mail já existente.	61
5.5	Caixa de criação de uma nova <i>thread</i>	61
5.6	Botão para mostrar ou esconder as palavras-chave de resumo para as <i>threads</i> de e-mail.	62
5.7	Listagem das palavras-chave de resumo de <i>threads</i> de e-mail.	62
5.8	Interface para filtragem de <i>threads</i> de e-mail entre trabalho, lazer ou indiferenciado.	62
5.9	Botão para criação de uma nova conversa.	62
5.10	Interface para mostrar informação dos participantes numa conversação.	63
5.11	Identificação para <i>threads</i> classificadas como trabalho.	63
5.12	Identificação para <i>threads</i> classificadas como lazer.	63
5.13	Interface da solução na fase de pré-classificação manual.	65
5.14	Interface da solução após a fase de pré-classificação manual.	65
5.15	Interface da solução com o filtro de lazer ativo mostrando os detalhes de um remetente.	66
5.16	Interface da solução com as palavras-chave de resumo ativas.	66
5.17	Interface da solução aquando da criação de uma nova <i>thread</i>	67
6.1	Opinião sobre a classificação a atribuir à seguinte mensagem de e-mail: “Tiago your work is awesome! We’ll buy your work by the value you want”.	73
6.2	Classificação automática atribuída pelo Spidious à mensagem anterior (A mensagem anterior faz parte de uma nova <i>thread</i> recebida na conversação com Tim Bake).	73
6.3	Nível de dificuldade sentido na execução da tarefa 1 - classificação de algumas <i>threads</i> de e-mail.	75
6.4	Nível de dificuldade sentido na execução da tarefa 2 - pesquisa de uma <i>thread</i>	75
6.5	Nível de dificuldade sentido na execução da tarefa 3 - resposta a uma mensagem de e-mail.	75

LISTA DE FIGURAS

6.6	Nível de dificuldade sentido na execução da tarefa 4 - obtenção das palavras-chave para resumo de <i>threads</i>	76
6.7	Nível de dificuldade sentido na execução da tarefa 5 - criação de uma nova <i>thread</i>	76
6.8	Nível de dificuldade sentido na execução da tarefa 6 - utilização do botão de criação de uma nova conversa.	77
6.9	Nível de dificuldade sentido na execução da tarefa 7 - aplicação de filtros à caixa de correio eletrónico.	77
6.10	Opinião sobre a utilidade da visualização orientada à conversa apresentada no Spidious.	78
6.11	Opinião sobre a utilidade da extração de palavras-chave, para resumo de <i>threads</i> , apresentada no Spidious.	78
6.12	Opinião sobre a utilidade da classificação de <i>threads</i> , em trabalho ou lazer, apresentada no Spidious.	79
6.13	Opinião sobre a utilidade da filtragem por mensagens indiferenciadas, de trabalho ou de lazer, apresentada no Spidious.	79
6.14	Opinião sobre o resultado final da visualização orientada à conversa apresentada no Spidious.	80
6.15	Opinião sobre o resultado final da extração de palavras-chave, para resumo de <i>threads</i> , apresentada no Spidious.	80
6.16	Opinião sobre o resultado final da classificação de <i>threads</i> , em trabalho ou lazer, apresentada no Spidious.	80
6.17	Opinião sobre o resultado final da filtragem por mensagens indiferenciadas, de trabalho ou de lazer apresentada no Spidious.	81
6.18	Opinião sobre quão fácil é o uso de toda a interface do Spidious, resultante da mudança de paradigma no e-mail, operada pelo mesmo.	81
6.19	Opinião sobre o aumento de eficiência de utilização que os novos métodos de visualização do Spidious conferem ao e-mail, face aos clientes habituais dos utilizadores.	82
6.20	Opinião sobre o aumento de eficiência de utilização que os algoritmos de aprendizagem do Spidious conferem ao e-mail, face aos clientes habituais dos utilizadores.	82
6.21	Opinião sobre o aumento de eficiência de utilização que a conjugação de uma nova visualização e de técnicas de aprendizagem num cliente de e-mail web (Spidious) poderiam conferir, face aos clientes habituais dos utilizadores.	83
6.22	Opinião sobre quão melhor os utilizadores acharam a experiência de utilização proporcionada pelo Spidious, em comparação com a fornecida pelos seus clientes de e-mail habituais.	83
6.23	Opinião dos utilizadores em relação a uma futura utilização do Spidious, caso o mesmo chegasse à fase de produção.	84
6.24	Opinião dos utilizadores em relação à disponibilidade de, no futuro, virem a pagar uma subscrição mensal para utilizarem o Spidious.	84
B.1	Tipo de cliente de e-mail utilizado.	116
B.2	Clientes de e-mail utilizados.	116
B.3	Local de consulta do e-mail.	117
B.4	Frequência de consulta do e-mail.	117
B.5	Tempo médio diário despendido no e-mail.	117
B.6	Mensagens de e-mail recebidas por dia.	118
B.7	Mensagens de e-mail enviadas por dia.	118
B.8	Número médio de mensagens não lidas na caixa de entrada.	118
B.9	Tipo de mensagens de e-mail deixadas na caixa de entrada.	119

LISTA DE FIGURAS

B.10 Número de pastas excluindo as existentes por defeito.	119
B.11 Método de pesquisa de mensagens mais utilizado.	119
B.12 Critério de pesquisa mais utilizado.	120
B.13 Tipo de notificações automáticas utilizadas.	120
B.14 Classificação das notificações automáticas quanto à sua intrusividade.	120
B.15 Frequência de utilização de uma mensagem já recebida para obter o e-mail do destinatário para uma nova mensagem.	121
B.16 Satisfação dos utilizadores em relação à atual interface dos clientes de e-mail. . .	121
B.17 Opinião dos utilizadores quanto à prejudicialidade da atual visualização do e-mail face à sua produtividade pessoal.	121
B.18 Opinião dos utilizadores quanto à vontade de terem um cliente de e-mail com uma visualização orientada ao remetente.	122
B.19 Opinião dos utilizadores quanto à vontade de terem um cliente de e-mail que permita uma resposta rápida tal como acontece num sistema de mensagens instantâneas.	122
B.20 Nível de satisfação com as funcionalidades de inteligência e aprendizagem já desenvolvidas para o e-mail.	123
B.21 Receção de e-mails sem interesse em ser lidos aquando da sua chegada.	123
B.22 Opinião sobre sobrecarga causada pela quantidade de mensagens de e-mail para processar.	123
B.23 Pretensão por um cliente de e-mail capaz de se adaptar ao perfil do utilizador. . .	124
B.24 Pretensão por perceber o conteúdo de um conjunto de mensagens antes mesmo de as ler através da visualização de meta-informação.	124
B.25 Opinião sobre aumento de produtividade e eficiência no uso do e-mail causado por uma classificação clara das <i>threads</i> em trabalho ou lazer.	124
B.26 Opinião sobre o sistema de classificação automática do Spidious para as restantes <i>threads</i> não classificadas pelo utilizador.	131
B.27 Opinião sobre a classificação a atribuir à seguinte mensagem de e-mail: “ <i>Tiago your work is awesome! We’ll buy your work by the value you want</i> ”.	131
B.28 Classificação automática atribuída pelo Spidious à mensagem anterior (A mensagem anterior faz parte de uma nova <i>thread</i> recebida na conversa com Tim Bake).	132
B.29 Opinião sobre quem iniciou a <i>thread</i> que fala sobre o campeonato português de futebol na conversa com o Tony Stacks.	132
B.30 Opinião sobre quantas mensagens compõe a <i>thread</i> referida anteriormente.	132
B.31 Opinião sobre se existem outras <i>threads</i> na conversa referida anteriormente. . .	133
B.32 Opinião sobre o número de mensagens com que a <i>thread</i> , referida anteriormente, ficou depois de efetuada uma nova resposta.	133
B.33 Gosto pela interface de resposta rápida a uma <i>thread</i>	133
B.34 Opinião sobre o número médio de palavras-chave geradas para cada <i>thread</i>	134
B.35 Opinião sobre se as palavras-chave geradas se encontram de acordo com o conteúdo da <i>thread</i> respetiva.	134
B.36 Gosto pela interface da funcionalidade de palavras-chave de resumo de <i>thread</i> . . .	134
B.37 Opinião sobre quão intuitiva se revelou a interface de palavras-chave de resumo de <i>threads</i>	135
B.38 Opinião sobre o número de <i>threads</i> com que a conversa com Tony Stacks ficou após a criação de uma nova <i>thread</i>	135
B.39 Gosto pela interface implementada para a criação de uma nova <i>thread</i> com um assunto diferente, dentro de uma conversa já criada.	135

LISTA DE FIGURAS

- B.40 Opinião sobre quão intuitivo se revelou o botão para criação de uma nova conversa. 136
- B.41 Opinião sobre quão intuitiva se revelou a interface que permite efetuar a filtragem na caixa de correio eletrónico. 136

LISTA DE FIGURAS

Lista de Tabelas

3.1	Três perfis de utilizadores na gestão do e-mail segundo S. Whittaker e C. Sidner. .	10
3.2	Descrição das diferenças entre os grupos de utilizadores encontrados por Jacek Gwizka.	12
3.3	Políticas de processamento de e-mail do artigo de Ashish Gupta, Ramesh Sharda e Robert Greve.	16
3.4	Análise comparativa dos trabalhos industriais.	36
4.1	Lista dos problemas encontrados pelos inquiridos nos clientes de e-mail atuais. .	50
4.2	Lista das funcionalidades pretendidas pelos inquiridos num novo cliente de e-mail.	51
6.1	Lista dos aspetos positivos e negativos, apontados pelos participantes do inquérito, do Spidious.	85

LISTA DE TABELAS

Abreviaturas e Símbolos

ACTA	<i>Activity-Centered Task Assistant</i>
AOL	<i>America Online</i>
CSS	<i>Cascading Style Sheets</i>
DNS	<i>Domain Name System</i>
EUA	<i>Estados Unidos da América</i>
HTML	<i>HyperText Markup Language</i>
IBM	<i>International Business Machines</i>
IMAP	<i>Internet Message Access Protocol</i>
JSON	<i>JavaScript Object Notation</i>
MIME	<i>Multipurpose Internet Mail Extension</i>
MTA	<i>Mail Transport Agent</i>
MUA	<i>Mail User Agent</i>
MVC	<i>Model-View-Controller</i>
PDF	<i>Portable Document Format</i>
POP	<i>Post Office Protocol</i>
RIPPER	<i>Repeated Incremental Pruning to Produce Error Reduction</i>
SMTP	<i>Simple Mail Transfer Protocol</i>
TCP	<i>Transmission Control Protocol</i>
TF-IDF	<i>Term Frequency–Inverse Document Frequency</i>
US-ASCII	<i>American Standard Code for Information Interchange</i>

Capítulo 1

Introdução

”O teu tempo é limitado, por isso não o desperdices a viver a vida de outra pessoa. Não te deixes armadilhar pelos dogmas - que é a mesma coisa que viver pelos resultados do que outras pessoas pensaram. Não deixes que o ruído das opiniões dos outros saia da tua própria voz interior. E, mais importante ainda, tem a coragem de seguir o teu coração e a tua intuição. Estes já sabem, de alguma forma, aquilo em que tu verdadeiramente te vais tornar. Tudo o resto é secundário.”

—Steve Jobs

As caixas de correio eletrónico continuam a ficar maiores, e fazer a gestão do e-mail torna-se cada vez mais difícil. Já se passaram quase vinte anos desde que Whittaker e Sidner [WS96] identificaram os maiores problemas do e-mail e a necessidade de uma mudança. No entanto, estudos posteriores, como o de Danyel Fisher et al. [FBGS06], referem que os problemas identificados mantêm-se praticamente todos por resolver, e os utilizadores continuam a sentir-se completamente sobrecarregados ao usar o e-mail.

O trabalho que se segue pretende abordar estes problemas amplamente estudados, mas praticamente pouco resolvidos. Serão analisados os trabalhos mais importantes já efetuados na área, quer a nível académico quer a nível industrial, e será também proposta uma solução que tentará resolver o problema, ou pelo menos atenuá-lo. É certo que o domínio de todo este problema é vasto e a sua solução claramente difícil. Contudo, difícil e complexo não são sinónimos, e remetendo-me para a citação de Steve Jobs no início deste capítulo, seguirei o meu coração e a minha intuição, porque de alguma forma eles já sabem aquilo que serei capaz de alcançar.

1.1 Contexto

O e-mail é uma das ferramentas de comunicação mais utilizadas no mundo, especialmente em ambiente empresarial. Estima-se que diariamente sejam enviados cerca de 196 biliões de e-mails e ainda que, até 2018, existirão aproximadamente 5.2 biliões de contas registadas [Sar14]. O e-mail foi projetado para ser usado em comunicação assíncrona, e sempre cumpriu esta tarefa com eficácia.

Introdução

No entanto, a verdade é que a massificação do seu uso fez com que as pessoas o começassem a usar para diferentes tipos de finalidades, as quais não foram pensadas originalmente. Estes acontecimentos levaram a que surgissem no e-mail diversos problemas, que estão devidamente identificados e que se arrastam há quase duas décadas a esta parte [WS96, WBM05, FBGS06], devidamente identificados na [Secção 5.1](#).

Os utilizadores perdem demasiado tempo e recursos, mais do que aqueles que deviam, ao utilizar o e-mail. Por cada interrupção, causada por um novo e-mail que chega, há um tempo de redução de produtividade do qual se demora a recuperar, e que causa grandes prejuízos a todas as partes [JDW01, JDW03a, JDW03b].

De facto, tem-se assistido a um grande interesse por parte de grandes empresas tecnológicas, em colocarem no mercado soluções próprias relacionadas com o e-mail. Isto só confirma a existência de problemas graves que permanecem por resolver. De entre soluções bastante conhecidas podemos destacar, por exemplo, o Inbox da Google, o ActiveInbox, o Mailbox da Dropbox ou até mesmo o Boomerang for Gmail da BayDIn. Mas será que já foi encontrada uma solução eficaz e eficiente para estes problemas, através das atuais soluções académicas ou industriais?

1.2 Motivação

Pelo que será possível perceber ao longo desta dissertação, existem diversas soluções disponíveis e muita investigação publicada nesta área. Contudo, percebe-se que os problemas não estão resolvidos. Não estavam em 1996 [WS96], não estavam em 2006 [FBGS06] e continuam a não estar quase vinte anos depois, em 2015.

O que é verdadeiramente perceptível é que os problemas principais estão claramente identificados. Na verdade, as soluções encontradas abordam de alguma forma alguns dos problemas. Mas o que fica bem patente é que nenhuma das encontradas consegue olhar para o problema de uma forma clara e fornecer uma solução suficientemente abrangente, eficaz e eficiente.

A verdadeira motivação para desenvolver este trabalho prende-se essencialmente com o que foi acabado de referir. As falhas apresentadas pelo e-mail representam um problema que afeta muita gente, que é bem compreendido mas que está mal resolvido.

1.3 Questões de Pesquisa

Este trabalho foi conduzido com o propósito de procurar responder a algumas questões fundamentais. A pesquisa, os inquéritos, o estudo com utilizadores efetuados, as publicações e trabalhos analisados que serão apresentados de seguida, assim como a solução proposta têm como objetivo perceber o seguinte:

1. Será que é possível melhorar a eficiência do utilizador no e-mail usando novos métodos e abordagens visuais?

2. Será que é possível melhorar a eficiência do utilizador no e-mail aplicando algoritmos de aprendizagem às contas de e-mail dos utilizadores?
3. Será que é possível melhorar a eficiência do utilizador no e-mail conjugando as abordagens anteriormente referidas num cliente de e-mail web?

Esta dissertação dará portanto resposta a estas perguntas, fazendo com que sejam mais perceptíveis todos os contornos envolvidos neste trabalho e nesta temática.

1.4 Objetivos e Contribuições Esperadas

O tema do e-mail não é de todo fácil, e portanto, os objetivos e as contribuições do trabalho final têm que ser bem definidas. Este trabalho tem como objetivo criar uma abordagem que permita resolver, ou pelo menos melhorar, os problemas verificados no e-mail, adaptada às exigências e padrões atuais. Por um lado, e representando o principal foco do trabalho, a solução abordará uma parte visual, diferente do que tem sido tentado até agora, que pretende permitir ao utilizador a gestão de um grande volume de mensagens. Por outro lado, de forma meramente complementar, a solução abordará também o uso de algoritmos de aprendizagem centrados em ações do utilizador, que permitem que em dado momento o utilizador apenas receba mensagens verdadeiramente importantes. De ter em conta que o uso de tais algoritmos tem como principal objetivo demonstrar a sua aplicabilidade e utilidade no e-mail.

Por tudo o que foi referido, as principais contribuições esperadas para esta dissertação concentrar-se-ão num cliente de e-mail web capaz de aumentar a eficiência do utilizador ao lidar com o e-mail, o que ao mesmo tempo vai de encontro aos resultados obtidos no [Capítulo 4](#). Para isso, a solução proposta por esta dissertação recorrerá a uma interface que se assemelhará à de uma aplicação de *chat*, que permitirá ao utilizador perceber o conteúdo de uma *thread*, antes mesmo de a ler, através do resumo da mesma em palavras-chave. A solução proposta recorrerá ainda a um conjunto de algoritmos de aprendizagem que, pelas ações do utilizador, perceberão se um dado e-mail se enquadra na categoria de *Trabalho* ou na categoria de *Lazer*.

1.5 Estrutura do Documento

Para além da introdução, que é o presente capítulo, esta dissertação contém mais seis capítulos. No [Capítulo 2](#), são apresentados alguns fundamentos que foram considerados pertinentes para melhor se entender esta dissertação. No [Capítulo 3](#), é descrito o estado da arte do e-mail e são apresentados trabalhos relacionados, sejam eles trabalhos académicos ou industriais. No [Capítulo 4](#), são detalhados os métodos utilizados para a elaboração de um questionário sobre o e-mail a membros da comunidade da Universidade do Porto (docentes e alunos). No mesmo capítulo são ainda apresentados os resultados do referido inquérito e a sua discussão. No [Capítulo 5](#), é apresentada a solução para o problema que esta dissertação pretende resolver. Para a determinação da

Introdução

solução final foram tidas em conta todas as pesquisas efetuadas, assim como os resultados provenientes do inquérito à comunidade da Universidade do Porto. No [Capítulo 6](#), são detalhados os resultados do teste efetuado a 20 utilizadores que serviu para validar a solução. No [Capítulo 7](#), por fim, são apresentadas as conclusões retiradas de todo o trabalho, as contribuições obtidas e uma pequena discussão do trabalho a ser realizado no futuro.

Capítulo 2

Fundamentos

Para que seja possível compreender melhor os temas abordados ao longo deste documento, é conveniente ter alguns conhecimentos prévios. Este capítulo pretende abordar alguns fundamentos sobre o e-mail, como por exemplo a arquitetura do serviço de e-mail e a definição e apresentação de alguns clientes de e-mail.

É esperado então que, no fim deste capítulo, seja possível compreender melhor o serviço de e-mail e assim, de forma rápida assimilar e perceber o trabalho, os problemas e as soluções existentes em volta desta temática.

2.1 Arquitetura do E-mail

Para melhor se entender o sistema de e-mail, a sua arquitetura tem de ser explicada. Desse modo, e começando pelo início, o e-mail foi desenhado para ser um sistema eletrónico assíncrono de troca de mensagens, seguindo algumas convenções e usando diversos protocolos. A arquitetura geral do sistema está descrita na [Figura 2.1](#).

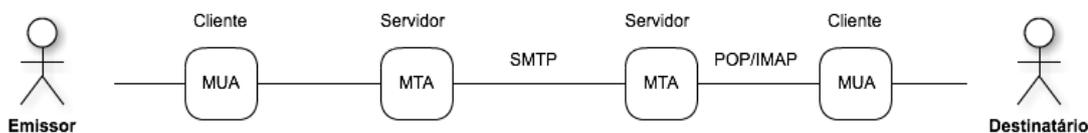


Figura 2.1: Imagem representativa da arquitetura base do sistema de e-mail.

O primeiro passo no uso do sistema é o de o utilizador escrever uma mensagem. O utilizador começa por fazê-lo usando um cliente de e-mail, também conhecido por MUA (*Mail User Agent*), sendo que depois de escrita, a mensagem é diretamente enviada para o servidor de e-mail do utilizador também conhecido por MTA (*Mail Transport Agent*). Depois deste passo, o MTA do remetente envia a mensagem ao MTA do destinatário utilizando a internet e um protocolo chamado SMTP (*Simple Mail Transfer Protocol*). A partir do momento em que a mensagem chega ao MTA do utilizador de destino, este pode efetuar a sua leitura utilizando dois protocolos: POP (*Post*

Office Protocol) e IMAP (*Internet Message Access Protocol*). O que difere estes dois protocolos é essencialmente a forma como eles gerem as mensagens de correio eletrónico. O protocolo POP descarrega todas as mensagens do servidor de e-mail para o computador do utilizador. O protocolo IMAP, por sua vez, permite ao utilizador ler as mensagens sem as remover do servidor de e-mail.

No entanto, falamos da arquitetura do e-mail e da troca de mensagens, mas não falamos diretamente da própria mensagem, que também é regida por normas. Uma mensagem de e-mail é geralmente dividida em duas partes principais: um cabeçalho e um corpo, sendo que, é o cabeçalho que guarda informação importante para o correto envio da mensagem, como o destinatário, ou a lista de destinatários, e o remetente. A norma que rege o cabeçalho e estrutura de uma mensagem de e-mail é a norma RFC822 [Cro82]. No entanto, esta norma mostrou-se insuficiente, uma vez que apenas permite caracteres US-ASCII. Para colmatar esta falha foi criado o protocolo MIME [BF92] (Multipurpose Internet Mail Extension), que fez com que, pudessem começar a ser enviados através do e-mail outros tipos de dados, como por exemplo imagens ou documentos.

Por fim, é importante referir que estas e outras informações, sobre toda a arquitetura do sistema de e-mail, podem ser consultadas e aprofundadas noutras fontes, como por exemplo livros e normas [PD11, Cro09, Cro82, BF92, BF93, FB96b, Moo93, FB96c, Moo96, FKP96, FB96a].

2.2 Cliente de E-mail

O cliente de e-mail, referido anteriormente como MUA, pode ser definido como uma aplicação que fornece ao utilizador uma interface para utilizar o e-mail: receber, ler, enviar, apagar e responder a mensagens. A gestão efetuada às mensagens recebidas depende dos protocolos usados pelo próprio cliente de e-mail. As mensagens recebidas podem ser diretamente armazenadas no computador do utilizador, caso o protocolo usado seja o POP, ou então armazenadas diretamente no servidor de e-mail, caso o protocolo usado seja o IMAP.

Por defeito os clientes de e-mail costumam utilizar três pastas diferentes: caixa de entrada, caixa de saída e lixo. No entanto, outras pastas podem ser criadas através destas aplicações para o utilizador poder gerir as suas mensagens como bem desejar.

Quanto à sua interface, um cliente de e-mail habitual tem uma visualização dividida em três painéis. Um primeiro painel, à esquerda, que lista as pastas do serviço de e-mail. Depois um painel central superior que lista todas as mensagens presentes na pasta selecionada. E por fim, um painel central inferior que mostra todo o conteúdo de uma dada mensagem.

É ainda importante referir que estas aplicações existem implementadas em três formas diferentes. Podem ser encontradas sob a forma de aplicações de *desktop* que têm que ser instaladas nos computadores dos utilizadores, como por exemplo o Thunderbird¹, sob a forma de clientes web, como por exemplo o Gmail², ou ainda sob a forma de aplicações móveis, como por exemplo o Mailbox³.

¹<https://www.mozilla.org/pt-PT/thunderbird/>

²<https://gmail.com>

³<http://www.mailboxapp.com/>

Capítulo 3

E-mail

Um tema como o do e-mail não é de todo fácil de abordar. Como se poderá evidenciar ao longo deste capítulo, têm sido realizadas várias tentativas nos últimos anos para agitar esta temática e apresentar algo comprovadamente melhor. Inesperadamente, ou não, o modelo convencional do cliente de email foi-se mantendo, na sua essência, inalterado.

Porém não é possível colocar de parte o que se foi fazendo de relevante nesta área, uma vez que, de uma maneira geral, as diferentes abordagens que foram divulgadas a público continham pormenores bem conseguidos, os quais devem ser objeto de estudo e revisão. Dessa forma, este capítulo pretende dar uma imagem geral da maioria dessas soluções e publicações apresentando o que, por um lado, elas tinham de melhor e, por outro, de pior. Esta análise será dividida fundamentalmente em duas partes: uma focando-se em trabalhos académicos (Interação Pessoa-Computador e Aprendizagem e Extração de Conhecimento), e outra focando-se em produtos e trabalhos industriais.

Para esta revisão de literatura foram usadas diferentes fontes de informação e termos de pesquisa. No que toca às fontes de informação foram usadas o *Google Scholar*¹, *ACM Digital Library*², *Springer*³, *IEEE Xplore Digital Library*⁴ e a *ProQuest*⁵. Já quanto aos termos utilizados, a pesquisa fundamentalmente foi feita com *email*, *email overload*, *email visualization*, *email mining*, *email classification*, *email summarization*, *email patterns*, *email usage*, *email behavior*, *email workflow*, *email management* e *email priority*. Como seria de esperar foram encontradas muitas publicações e muito material passível de ser objeto de estudo. Contudo, foi efetuada uma seleção com base nas citações das publicações encontradas, e o seu consequente grau de importância e confiança dentro da comunidade científica, e também a sua dispersão temporal na tentativa de tornar patente, nesta revisão de literatura, uma evolução. Ainda de referir que, ao longo deste capítulo a análise efetuada dentro de cada secção segue a ordem cronológica das publicações.

¹<https://scholar.google.pt>

²<http://dl.acm.org>

³<http://link.springer.com>

⁴<http://ieeexplore.ieee.org/search/advsearch.jsp>

⁵<http://search.proquest.com/advanced>

3.1 Interação Pessoa-Computador

A Interação Pessoa-Computador tem assumido um papel fundamental na forma como, atualmente, o *software* é desenvolvido. Por definição a Interação Pessoa-Computador estuda de que modo a tecnologia informática influencia as atividades e o trabalho humano [Dix09]. As suas áreas de intervenção compreendem simples computadores pessoais, dispositivos móveis, sistemas de navegação ou até mesmo sistemas embebidos. Desta forma, e sendo o e-mail uma aplicação intensamente utilizada, a Interação Pessoa-Computador, em particular a sua disciplina de *design* centrado no utilizador, empenhou-se em estudar esta aplicação de forma aprofundada, e até mesmo em propor novas soluções com o objetivo de resolver problemas já identificados [WBM05, WS96].

Por isso, será agora realizada uma revisão ao que de relevante foi publicado envolvendo o e-mail e a Interação Pessoa-Computador, sendo que se dividiu a análise em três temas: Padrões de Utilização, Fluxo de Trabalho e Métodos de Visualização.

3.1.1 Padrões de Utilização

Os problemas no e-mail foram desde muito cedo notados. Um dos primeiros a percebê-lo foi, muito provavelmente, Peter J. Denning [Den82] que na sua publicação reconhece que há duas etapas na geração de informação: a fase de preparação do documento, e a fase de disseminação do documento. Na sua opinião da época, o e-mail apoiava muito todos aqueles que geravam informação na fase disseminação, mas ao mesmo tempo já lhe anotava muitas lacunas dado que o mesmo não tinha bons meios para prevenir a receção não desejada. Para tentar melhorar, ou até mesmo resolver a situação, ele sugeriu que tinham de ser criados meios que permitissem a existência de caminhos especiais pelos quais mensagens pessoais, urgentes ou certificadas pudessem chegar, sendo que todos os outros caminhos teriam de ser filtrados.

A publicação referida no parágrafo anterior foi só o início de um longo caminho, isto porque depois dela surgiram muitos outros estudos que analisaram bastantes aspetos, no que aos padrões de utilização do e-mail diz respeito. Alguns anos depois Wendy Mackay [Mac88] publicou um estudo (baseado no sistema de lente de informação publicado por Malone et al. [MGT⁺87]) com o qual avaliou os padrões de utilização do e-mail por parte dos utilizadores, tornando-se desse modo um dos primeiros a fazê-lo. Desse modo, conseguiu aferir que o e-mail é maioritariamente utilizado para três fins de gestão diferentes: gestão de informação, gestão do tempo e gestão de tarefas, fins esses que estão associados a três grupos diferentes nos quais normalmente os utilizadores se inserem: *prioritizers*, *archivers* e *performers*. Os *prioritizers* são os que usam o e-mail como uma ferramenta para gestão de tempo e que, na globalidade, estão mais interessados em identificar e priorizar mensagens importantes. Os *archivers*, que usam o e-mail como uma ferramenta para gestão de informação, estão principalmente interessados em pesquisar e obter mensagens de e-mail. O último grupo de utilizadores identificados, os *performers* usam o e-mail com um intuito de gestão de tarefas uma vez que normalmente identificam e atribuem tarefas aqueles que acham serem mais capazes de as realizar. De referir ainda que Mackay classificou também os utilizadores em três categorias no que diz respeito ao momento em que eles têm de

lidar com todos os seus emails: *overwhelmed*, *on the edge* e *ok*. Por fim, nas suas conclusões, Wendy Mackay refere que ver o e-mail como uma ferramenta de suporte para diferentes tipos de trabalhos e pessoas pode ajudar, seriamente, na criação de sistemas de maior sucesso. Ou seja, os *designers*, ao invés de procurarem um subconjunto de funções, devem pensar em diversidade criando primitivas poderosas que forneçam flexibilidade, e que possam tornar o correio eletrónico numa ferramenta sofisticada que represente mais que um sistema de comunicação.

O percurso da investigação na área não ficou por aqui e Steve Whittaker e Candace Sidner [WS96] publicaram um estudo no qual realizaram a análise quantitativa às caixas de correio eletrónico de vinte utilizadores, a par de trinta e quatro horas de entrevistas para tentarem cobrir a lacuna de informação existente sobre o modo de como as pessoas organizam e gerem grandes quantidades de informação e mensagens. Eles afirmam que o e-mail foi pensado originalmente para comunicação assíncrona mas que, como a sua análise demonstra, evoluiu até ao ponto de ser utilizado para diferentes propósitos como armazenamento e envio de documentos, delegação e verificação de tarefas, armazenamento de contactos e nomes, envio de lembretes, pedidos de assistência, agendamento e ainda para gerir questões de suporte técnico. Como resultado de todos estes propósitos diferentes identificaram ainda as três funções principais do e-mail como sendo a gestão de tarefas, o armazenamento pessoal e a comunicação assíncrona. De acordo com os dados recolhidos puderam ainda constatar que os utilizadores estavam interessados em poder agrupar e visualizar de forma conjunta mensagens semanticamente relacionadas, e que o modelo de ação de toque único estava ultrapassado uma vez que, assume que o utilizador, para mensagens recebidas, apenas quer proceder a ações de resposta, eliminação ou reencaminhamento. O que acaba por acontecer, segundo os autores, é que na prática existem mensagens que não podem ser processadas quando chegam, e acabam por permanecer na caixa de entrada para serem processadas posteriormente. Foram ainda identificadas três estratégias, comuns entre os utilizadores, para lidar com a sobrecarga de mensagens no correio eletrónico: *No filers*, *Frequent filers* e *Spring cleaners*. Os *No filers* são caracterizados por em regra não usarem as pastas do e-mail mas confiarem e usarem a pesquisa para encontrarem a informação que necessitam. Os *Frequent filers* fazem tentativas significativas para minimizar o número de mensagens na sua caixa de entrada, procedendo a passagens diárias pelas suas caixas de correio apagando ou arquivando as mensagens. Os utilizadores que fazem parte do último perfil identificado, os *Spring cleaners*, são caracterizados por lidarem com a sobrecarga das suas caixas de correio procedendo a fortes limpezas, normalmente em períodos que variam entre um e três meses. Usam ainda intensivamente pastas, mesmo que de forma ineficiente uma vez que, a maior parte das mesmas contém três mensagens ou menos. Os resultados empíricos completos deste estudo estão listados na [Tabela 3.1](#) que se segue, sendo os mesmos bastante interessantes e totalmente elucidativos.

Uns anos mais tarde Ducheneaut e Bellotti [DB01] publicaram um estudo onde afirmam que o e-mail se tornou mais um hábito que uma aplicação, e, sobretudo, devido à maneira como é usado um hábito mau, apelidando-o mesmo de *serial-killer*. Os autores apresentam as suas descobertas de quatro meses de trabalho conduzido em ambiente empresarial e referem ainda que confirmam as suas suspeitas de que os utilizadores usavam o e-mail como uma ferramenta de gestão de in-

E-mail

Strategy	N	# Inbox items	Total # items	Inbox as % of total mailbox	Old inbox items (%inbox> 3 month old)	New inbox items (%inbox< 1 month old)	# Inbox conversation threads	# Folders	Failder folders (# folders with <3 items)	Daily # messages received
No filers	6	3093.5	3271.1	95.25	51.58	11.78	287.5	11.33	4.5	58
Spring Cleaners	7	1492.29	2818.71	51.02	40.15	24.22	257.86	61.43	30.85	45.71
Frequent filers	5	43.4	1062.2	4.96	4.75	90.34	3.6	70.6	16.6	42

Tabela 3.1: Três perfis de utilizadores na gestão do e-mail segundo S. Whittaker e C. Sidner [WS96].

formação pessoal. As suas conclusões basearam-se principalmente na observação de uma grande preferência que os mesmos tinham por incorporar no e-mail o seu trabalho diário, fazendo desta junção o seu ambiente de trabalho preferencial. No seguimento dessa descoberta puderam de facto afirmar que o e-mail se encontrava a ser usado para mais funções do que aquelas para as quais foi desenhado. Contudo, eles ressaltaram também que esse facto dependia em muito de fatores como a profissão e as atividades desempenhadas pelo utilizador, assim como pela natureza do seu local de trabalho. Por fim, deixaram bem patente que o seu estudo poderá ter um grande impacto em todos aqueles que pretendam construir a próxima geração de clientes de e-mail, justificando-o com dois argumentos:

Primeiro Demonstraram a possibilidade de poderem serem feitas melhorias significativas a nível da interface do utilizador: para melhor suportar o uso do e-mail como uma ferramenta de gestão de informação pessoal, a organização das pastas tem de ser mais flexível; como maioritariamente os utilizadores trabalham com informação pode ser desenvolvido um sistema que mostre os itens recentemente acedidos; a gestão de tarefas e lembretes tem de ser incluída nos clientes de e-mail.

Segundo Devido ao facto de o e-mail ser altamente usado para efetuar troca de documentos têm de ser incorporadas algumas funcionalidades de gestão de documentos sendo que, uma das mais desejadas é um sistema de revisão de versões.

Em jeito de conclusão, os autores deixam ainda a pergunta se seria possível criar um cliente de e-mail no qual a sua interface fosse diferente, ou se adaptasse, a diferentes tipos de padrões de utilização, onde, por exemplo, fosse apresentada uma interface com planeamento e controlo de tarefas e atividades para um gestor, e onde fosse apresentada uma interface com funcionalidades de gestão de ficheiros e documentos a alguém que precisasse de um ambiente altamente colaborativo.

Dando seguimento ao estudo até agora apresentado, encontrou-se uma publicação, do mesmo ano que o trabalho falado no parágrafo anterior, efetuada por Venolia et al. onde se constata que foi desenvolvido um modelo conceptual das atividades dos utilizadores no correio eletrónico, com base num estudo no terreno, efetuado na empresa dos autores. As atividades encontradas pelo estudo foram as de *Flow*, *Triage*, *Task Management*, *Archive* e *Retrieving*. A atividade de *Flow* é descrita como o processo de acompanhar, de forma contínua, o fluxo das mensagens recebidas. A atividade de *Triage* é caracterizada como um processo de análise prévia dos e-mails com o

objetivo de saber como processá-lo, seja por uma estratégia de processamento em série ou por uma estratégia de processamento por prioridades. Já a atividade de *Task Management* caracteriza-se pelo uso das mensagens de correio eletrônico como listas de tarefas. A atividade de *Archive* foi definida como sendo um processo no qual o utilizador arquiva mensagens para futuramente usar no conteúdo das suas mensagens ou anexos. Em último lugar, a atividade de *Retrieving* foi caracterizada como sendo o processo de obter as mensagens antigas do arquivo. É ainda de referir que, ao longo do artigo, para cada atividade identificada foi feita uma análise entre as necessidades reais do utilizador e aquilo que um cliente de e-mail comercial oferece e quais as alterações que foram operadas, nos anos transatos, aos problemas já identificados, tendo os autores chegado à conclusão que maioritariamente, e de forma espantosa, os problemas se mantinham inalterados.

Por sua vez, num outro trabalho, Fisher e Moody [MMF⁺02], realizaram um estudo, com dados de caixas de correio eletrônico provenientes de setenta e seis utilizadores, com o objetivo de estudarem informações estruturais e comportamentais. Depois de colocadas várias hipóteses, que tinham como objetivo perceber a evolução ocorrida desde as constatações de outros trabalhos anteriores, chegaram à conclusão que o comportamento organizacional dos utilizadores não mudou, de forma relevante, desde o estudo de Whittaker e Sidner [WS96] e que as *threads* continuavam a envergar um grau de importância elevado nas caixas de correio dos utilizadores (cerca de 30% das mensagens apresentavam-se na forma de *threads*). De referir ainda que, nesta temática das *threads*, um estudo que sucedeu o de Fisher e Moody em dois anos, apurou, depois de uma análise a 42,000 mensagens de correio eletrônico, que 62% das mensagens diziam respeito a *threads* [Ker03]. A discrepância tão elevada em dois estudos que diferem pouquíssimo em termos temporais, pode ser explicada pelo facto de este segundo ter tido em conta as mensagens raiz em falta, tenham sido as mesmas apagadas ou simplesmente ficado esquecidas aquando de um envio.

No seu estudo sobre os perfis dos utilizadores no e-mail Jacek Gwizka [Gwi04], elaborado através de um questionário que contou com vinte e quatro participantes, concluiu que existem dois tipos de utilizadores diferentes: os *cleaners* e os *keepers*. Os primeiros são caracterizados por mover informação de futuras tarefas ou trabalhos da caixa de correio, enquanto os segundos são caracterizados por deixar essa informação de futuras tarefas ou trabalhos armazenadas nas caixas de correio eletrônico. Apresenta-se de seguida a Tabela 3.2 retirada do referido estudo que descreve melhor as características de cada grupo.

Os autores referem ainda que as diferenças encontradas entre os dois grupos referidos podem ser explicadas pelos diferentes níveis de experiência de utilização de e-mail verificados. Constataram ainda que, por um lado, não encontraram nenhuma correlação entre o comportamento para com o correio eletrônico e o género ou a organização das secretárias de trabalho dos utilizadores mas que, por outro lado, detetaram que o comportamento para com a caixa de correio eletrônico é afetada por fatores como a experiência prévia, o estilo cognitivo, e a função profissional.

Até ao momento, os trabalhos que foram apresentados, na sua maioria, baseavam o seu foco de estudo nas próprias mensagens de e-mail. Contudo, um estudo elaborado por Zhixian Yi [Yi07], com enfoque nos EUA (Estados Unidos da América), tentou determinar quem nos EUA utiliza mais o correio eletrônico. Para isso foi usada uma correlação bivariada e uma regressão linear

Email Habit Variables	The Cleaners (Cluster 1)	The Keepers (Cluster 2)
When email is read	read email at specific times	read email all the time
Email interrupts other tasks	email does NOT interrupt other tasks	email interrupts other tasks
Uses search in email	do NOT search in email	search in email
Keeps events in email	do NOT keep events	keep events
Keeps to-do's in email	do NOT' keep to-do's	keep to-do's
Emails self-reminders	send self-reminding email messages	do NOT send self-reminding email messages

Tabela 3.2: Descrição das diferenças entre os grupos de utilizadores encontrados por Jacek Gwizka [Gwi04].

múltipla entre variáveis sócio-demográficas tais como a idade, região e variáveis sócio-económicas tais como a educação, o salário e ainda a variável do tempo disponível para ter o trabalho feito. As principais conclusões do estudo foram:

- Os participantes no estudo com maiores níveis de educação, maior salário ou com tempo suficiente para fazer os seus trabalhos são os mais propensos a usarem o e-mail mais horas por semana;
- Os participantes no estudo de maior idade mostraram ser menos propensos a serem os que usam o e-mail mais horas por semana;
- O género, a raça e o estado civil não fazem qualquer diferença.

Por fim, o último trabalho publicado que irá ser analisado nesta [Subsecção 3.1.1](#) foi elaborado por Karagiannis e Vojnovic [KV09] e compreende um estudo dos perfis comportamentais que caracterizam os utilizadores, as suas ações ao processar os e-mails e algumas propriedades dos e-mails trocados. A análise foi feita com base nas mensagens de correio eletrónico trocadas numa empresa multinacional de larga escala (cerca de 315 milhões de mensagens analisadas), com mais de 100 000 funcionários dispersos por vários países . Os autores focaram-se em perceber principalmente quais os fatores que influenciam a resposta às mensagens de e-mail . Das conclusões do estudo verificou-se que os dois fatores que mais influenciam a resposta a um e-mail são: o tamanho da lista dos destinatários e o número de e-mails trocados entre o remetente e o destinatário.

3.1.2 Fluxo de Trabalho

Depois de serem analisados os padrões utilização no e-mail na [Subsecção 3.1.1](#) é importante analisarmos os estudos realizados sobre os fluxos de trabalho existentes e recomendados para o

e-mail. Um dos primeiros estudos a abordar esta temática, e consequentemente a sugerir algumas dicas de como o melhorar, foi publicado por Jackson Thomas, Ray Dawson e Darren Wilson [JDW01]. No seu trabalho eles constaram que as pessoas permitem ser interrompidas pelo e-mail quase tão frequentemente como por chamadas telefónicas (em média a cada cinco minutos) e que, a reação normal à receção de um e-mail era uma resposta o mais rápida quanto possível (em média em seis segundos), ao invés de apenas responder na altura mais apropriada. Contudo verificaram ainda que, mesmo sendo o e-mail menos disruptivo que uma chamada telefónica ainda continua a interferir, mais do que devia, no fluxo de trabalho das pessoas, uma vez que devido às suas más práticas de uso, por cada vez que são interrompidas demoram cerca de sessenta e quatro segundos a recuperarem o seu ritmo de trabalho. Desse modo, os autores da publicação tentaram propor uma solução. Analisando um conjunto de dados os autores criaram um conjunto de recomendações e boas práticas no uso do correio eletrónico com o objetivo de aumentarem a eficiência e a produtividade dos utilizadores:

1. Reduzir a proeminência das interrupções desligando as notificações visuais e sonoras de alerta para a chegada de uma nova mensagem de e-mail.
2. Restringir o uso das mensagens de e-mail para grandes listas de destinatários, em particular a funcionalidade de responder a todos. O uso de grupos de e-mail mais reduzidos e mais focados pode ajudar a resolver o problema.
3. Configurar a aplicação de e-mail para mostrar, ainda na caixa de entrada, o remetente, o assunto e três linhas do conteúdo da mensagem para o utilizador poder rapidamente determinar se o e-mail requer uma atenção imediata.
4. Configurar a aplicação de e-mail para apenas procurar por novas mensagens de e-mail num período nunca inferior a quarenta e cinco minutos.
5. Num ambiente empresarial, oferecer sessões de formação a todos os funcionários sobre como utilizar o e-mail, em áreas como: definir a prioridade de uma mensagem, manter e guardar mensagens obedecendo a certas regras, utilizar eficientemente grupos de utilizadores, listas de contacto e ainda construir mensagens melhor estruturadas.

Também segundo estes autores, o estudo demonstrou a importância da medição dos processos de comunicação, uma vez que depois de ter sido feito no e-mail mostrou resultados surpreendentes, que não estavam a ser tidos em conta, e que por conseguinte levaram à criação de boas práticas que pretendem melhorar o problema, e que podem ser implementados em qualquer grupo de trabalho.

Um outro trabalho publicado no mesmo ano por J.J. Cadiz et al. [CDGV01] aborda exatamente o mesmo problema. Através dos resultados de um estudo realizado numa empresa, os autores construíram um modelo conceptual e identificaram as cinco maior atividades no que diz respeito à forma como os utilizadores utilizam o e-mail:

Flow: Uma vez que as pessoas continuam a trabalhar noutras tarefas variadas, elas pretendem lidar com o fluxo de entrada de mensagens, da melhor forma possível, à medida em que elas

vão chegando. Para intervir e melhorar esta atividade os autores sugerem que os clientes de e-mail gerem notificações visuais com informação suficiente que permita ao utilizador decidir se pretende, ou não, efetuar algum tipo de ação sobre a referida mensagem. No entanto, referem ainda que deve ser usado um método de priorização já que algumas mensagens, definitivamente, não precisam de ser processadas de imediato.

Triage: Depois de as pessoas se ausentarem do seu correio eletrónico por um dado período de tempo, elas precisam de lidar com todos os e-mails acumulados enquanto estiveram ausentes. Relativamente a esta atividade, os autores fazem algumas sugestões de melhoria, como por exemplo dividir a caixa de entrada em diferentes categorias, dentro das quais as mensagens aparecem organizadas por ordem de importância. Sugerem ainda a implementação de threads de conversação uma vez que permitem reduzir o número de mensagens dentro das caixas de correio eletrónico, assim como permitem visualizar as mensagens dentro do seu contexto.

Task Management: As pessoas usam o e-mail frequentemente para se lembrarem do que precisam de fazer, e para obterem ajuda a completar as tarefas. Para melhorar esta atividade complexa, os autores sugerem que se criem formas de poder marcar as mensagens que estão pendentes, sugerindo as seguintes marcas: *Pending - Action required*; *Pending - To be read*; *Pending - Need to respond* and *Pending - Unspecified*.

Archive: O armazenamento de e-mails é uma atividade comum, uma vez que as pessoas guardam mensagens para mais tarde as puderem referir ou usar. Para melhorar esta atividade foi sugerido pelos autores que a uma mensagem pudessem ser associados um número arbitrário de etiquetas e que as etiquetas fossem associadas a *threads* e não a mensagens, reduzindo assim a quantidade de trabalho necessário no processo de arquivamento.

Retrieving: Mais tarde, depois de terem armazenado as mensagens, as pessoas precisam de uma maneira de as recuperar. As sugestões dos autores para melhorar esta atividade prendem-se sobretudo com a criação de uma boa interface que forneça facilidade e ferramentas para refinar as pesquisas. Os filtros principais nas pesquisas seriam a etiqueta, o remetente, a lista de discussão, a data e os anexos. Sugeriram também ainda que, o resultado das pesquisas deve ser sempre retornado no contexto de uma *thread*, ao invés de ser retornado no contexto da mensagem.

Depois de apresentarem e referirem as atividades apresentadas anteriormente, os autores referem que na literatura encontrada anteriormente, tem sido prestada pouca atenção às duas primeiras atividades e que as mesmas são de grande importância. Por fim, propõe ainda a implementação de um cliente de e-mail que implemente *threads* de conversação, uma vez que as mesmas são um tema recorrente em discussões sobre melhoramentos a aplicar a interfaces com o utilizador, e no e-mail, comprovadamente funcionam bem.

Um outro trabalho que se foca nesta temática do fluxo de trabalho foi publicado por Joshua Tyler e John Tang [TT03] e foca-se numa análise dos ritmos da atividade no e-mail com o objetivo

de poder ajudar os utilizadores a coordenar a sua correspondência eletrónica. Os autores partiram da premissa de que o e-mail é claramente uma ferramenta crucial para as empresas, e que perceber os diferentes tipos de ritmos que orientam o seu uso poderá ajudar a identificar padrões de desenho capazes de melhorar a eficiência do serviço. O seu estudo baseou-se em questões como:

- Como é que os utilizadores decidem quando ler e responder às mensagens?
- Como é que os utilizadores formam expectativas acerca de quanto tempo vai demorar até que os outros lhes respondam?
- Como é que estas expectativas afetam os comportamentos no e-mail?

Os autores concluíram que a estrutura do correio eletrónico na época fornecia poucas ferramentas para perceber quando os utilizadores esperavam uma resposta ou quando seria uma boa altura para a resposta ser enviada. Da sua investigação acerca do tema surgiu um padrão comum entre os inquéritos realizados. Na comunicação de e-mail identificaram que há duas fases distintas: uma fase que representa a expectativa de resposta, e outra que representa o período de quebra em que o utilizador tenta de novo obter uma resposta. Estas duas fases de diferente comunicação podem ser observadas melhor na [Figura 3.1](#).

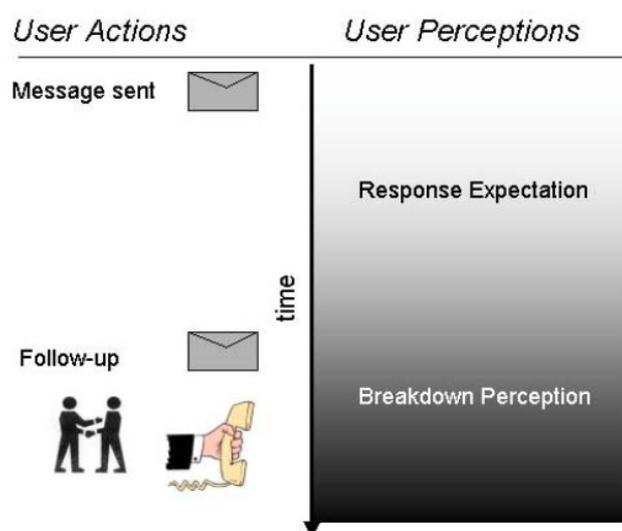


Figura 3.1: Representação das duas fases presentes na comunicação via e-mail: da expectativa à quebra. Imagem retirada do artigo de Joshua Tyler e John Tang [TT03].

É ainda referido que a fronteira entre as duas fases, a de expectativa de resposta e a de quebra, não é sempre a mesma, uma vez que varia de e-mail para e-mail. Os fatores de influência verificados são vastos, sendo que foi evidenciado que fatores como o recetor, a localização do recetor, a urgência do assunto e uma comunicação combinada com o *voicemail* afetam altamente essa variância. Com a referida investigação é esperada a construção de sistemas de correio eletrónico que

tenham em conta este fator, e esta linha temporal encontrada pelos autores, e que dessa forma seja possível alterar o fluxo de trabalho do e-mail para melhor.

Continuando com a análise, encontra-se um outro estudo de Ashish Gupta, Ramesh Sharda e Robert Greve [GSG10] que tenta, através de uma análise de dados recolhidos através de uma simulação rigorosa, identificar qual a melhor política para processar mensagens de correio eletrónico tendo em vista minimizar o tempo perdido. É referido ainda que estudos anteriores sugeriam que a maneira mais eficiente de processar as mensagens de e-mail seria em turnos de 45 minutos. O estudo é efetuado às políticas referidas na seguinte [Tabela 3.3](#):

Processing strategies	Descriptions
Continuous attention (C)	This processing strategy requires processing email as they arrive (giving first priority to email).
Scheduled attention-1 (C1)	This processing strategy requires holding email hours once daily, every morning.
Scheduled attention-2 (C2)	This processing strategy requires holding email hours twice daily.
Scheduled attention-2P	This processing strategy requires holding email hours twice daily, during two peak email arrival periods.
Scheduled attention-4 (C4)	This processing strategy requires holding email hours four times daily.
Scheduled attention-4P	This processing strategy requires holding email hours four times daily, during four peak arrival periods.
Scheduled attention-6 (C6)	This processing strategy requires holding email hours six times daily.
Jackson attention (C8)	

Tabela 3.3: Políticas de processamento de e-mail do artigo de Ashish Gupta, Ramesh Sharda e Robert Greve [GSG10].

As conclusões deste estudo apontam para uma recomendação essencial, a de o processamento de e-mails de forma contínua ou com alta frequência não ser de todo a melhor política para reduzir o efeito de interrupção causado pelo e-mail. O estudo mostra ainda que verificar a caixa de correio eletrónico duas a quatro vezes por dia é a melhor política para ambientes de trabalho, uma vez que a referida política tende a reduzir o sentimento de sobrecarga e o tempo perdido pelas interrupções do e-mail ao mesmo tempo que consegue também um ótimo balanceamento entre a conclusão de tarefas e o tempo de resposta às mensagens de correio eletrónico. Os autores descobriram ainda que é uma boa política não ter poucas (política C1) ou demasiadas (políticas C8 e C) horas prioritárias para lidar com o e-mail, uma vez que o número ótimo se encontra algures no meio (políticas C2 e C4).

O último trabalho a ser analisado nesta temática é um estudo recente publicado por Milos Vacek [Vac14]. A publicação assenta numa investigação levada a cabo em 2012 numa universidade da República Checa (*University of Hradec Králové*) e apresenta os requisitos dos utilizadores para comunicação eletrónica ao mesmo tempo que descreve formalmente um processo para lidar com o e-mail. O processo descrito foi representado pelo autor sob a forma de um diagrama, que pode ser consultado na [Figura 3.2](#), e todo ele será descrito de seguida:

Primeira Leitura O utilizador rapidamente identifica o conteúdo da mensagem e decide sobre futuras ações sem nunca o abrir. Idealmente este seria um processo que não deveria requerer qualquer tipo de ação humana.

Tomada de Decisão Depois da atividade anterior o utilizador decide o que fazer em seguida com a mensagem de e-mail: apaga o que não é importante, e guarda o que é importante para mais

E-mail

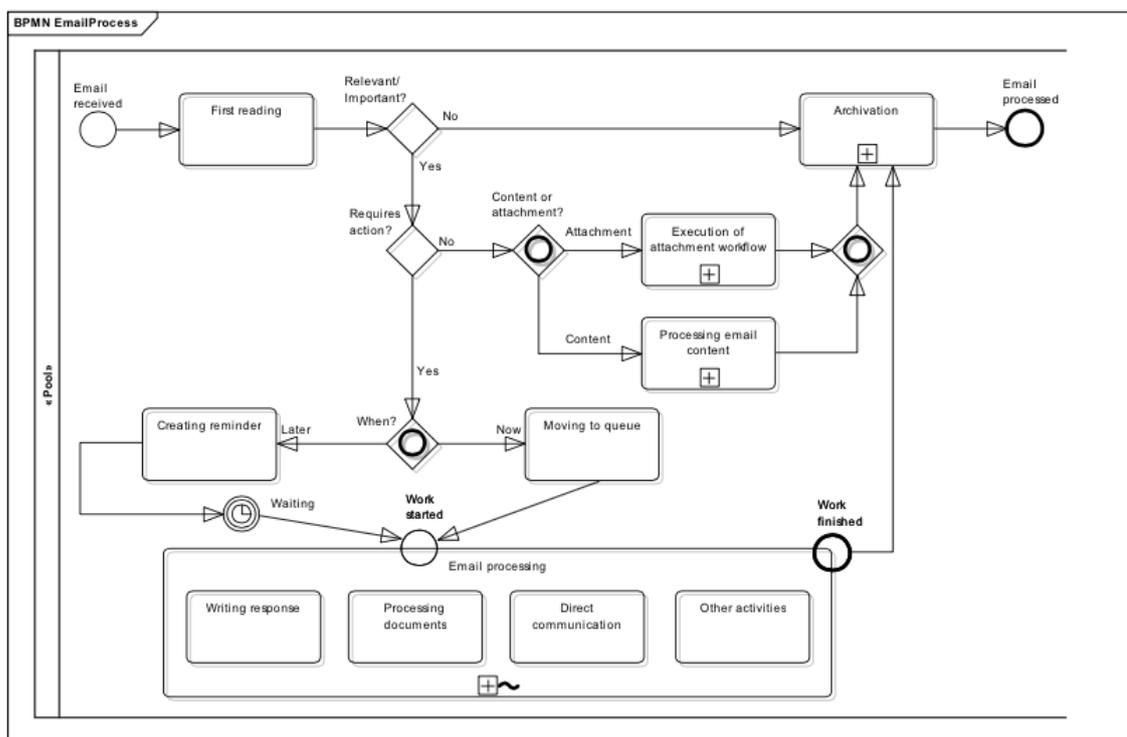


Figura 3.2: Processo sugerido para lidar com o e-mail segundo Milos Vacek [Vac14].

tarde processar de acordo com a importância. Segundo o autor, este seria um processo sério a ser automatizado.

Arquivamento Nesta fase os e-mails são ordenados de acordo com vários fatores. As mensagens importantes são movidas para pastas de tarefas e depois de completas para a pasta das tarefas completas. O autor sugere que às pastas de arquivo poderiam ser associadas funcionalidades de limpeza automática.

Execução do Fluxo de Trabalho para Anexos Esta atividade é gerida de diferentes formas por diferentes utilizadores. Segundo o autor um grande melhoramento seria a implementação de uma funcionalidade que permita ao e-mail reconhecer automaticamente o formato dos anexos associados e consequentemente oferecer um fluxo de trabalho pré-definido para cada tipo.

Processamento do Conteúdo dos E-Mails Esta atividade contempla diversas acções dependendo do conteúdo, mas, segundo o autor, o uso apropriado de categorias para as mensagens de e-mail poderia ser uma solução eficaz.

Mover para uma Fila O autor refere que uma lista de tarefas integrada pode priorizar mensagens e fazer com que os utilizadores possam lidar com o e-mail em determinados tempos reservados.

Criar um Lembrete Nesta atividade o autor refere que muitos clientes de e-mail não permitem criar lembretes para uma determinada mensagem, o que faz com o que os utilizadores tenham de utilizar ferramentas externas para o conseguirem fazer. Ele refere que esta é uma funcionalidade essencial que pode melhorar significativamente o fluxo de trabalho no e-mail.

Processamento do E-Mail Nesta atividade, segundo o autor, encontra-se uma quantidade de sub-atividades enorme, e que cada uma delas, individualmente, pode ser resolvida de forma diferente. O autor sugere que antes de um utilizador escrever uma resposta à mensagem de e-mail deveria verificar se o destinatário se encontra disponível para conversar em alguma ferramenta de *chat*, ou então efetuar uma chamada telefónica para o mesmo.

Por fim o autor refere que a sobrecarga originada pelo e-mail não é apenas causada por aspetos psicológicos mas também por limitações tecnológicas que podem ser resolvidas caso a sua solução seja implementada e posta em prática. Como é referido também pelo autor, o correio eletrónico não é apenas um serviço extra que requer alguma atenção por parte dos seus utilizadores, mas sim um meio de comunicação padrão que faz parte das tarefas diárias.

3.1.3 Métodos de Visualização

Após termos anteriormente analisado aquilo que foi feito à cerca dos perfis de utilização e sobre os fluxos de trabalho no e-mail, faz todo o sentido terminarmos a nossa análise acerca do que foi publicado dentro da área da interação pessoa-computador revendo agora as publicações acerca de métodos de visualização relacionados com o correio eletrónico.

O primeiro trabalho encontrado na pesquisa efetuada data de 1997, foi publicado por Kelvin Yiu et al. [YBSL97] e diz respeito a um projeto de seu nome TimeStore. O TimeStore é uma interface para o e-mail que permite uma gestão de tarefas e uma visualização temporal das mensagens. Esta interface organiza as mensagens pela sua data e hora de chegada e distribui-as como pontos num espaço bidimensional de tempo e remetentes. Permite também o uso de pastas de correio eletrónico dinâmicas que por exemplo servem para listar as mensagens novas ou não lidas. O TimeStore permite ainda a criação de tarefas a partir da janela de leitura da mensagem, o que torna esta funcionalidade bastante atrativa. A interface do TimeStore e as características referidas estão bem visíveis na [Figura 3.3](#).

A próxima publicação em análise é da autoria de Judith Donath, Karrie Karahalios e Fernanda Viegas [DKV99] e diz respeito a uma ferramenta chamada Loom. O Loom não é diretamente um cliente de e-mail mas sim uma ferramenta de visualização para as conversações dos grupos *Usenet*. No entanto, as conversas destes grupos eram feitas em forma de *thread* e os clientes de e-mail de hoje em dia, por defeito, têm este tipo de visualização implementada. Esta ferramenta representava cada mensagem num espaço bidimensional onde a variável tempo era representada no eixo dos x e a variável remetente era representada no eixo dos y. As mensagens eram ligadas por linhas que representavam as diferentes interações dentro e fora de uma mesma *thread*. A interface da ferramenta pode ser vista na [Figura 3.4](#).

E-mail

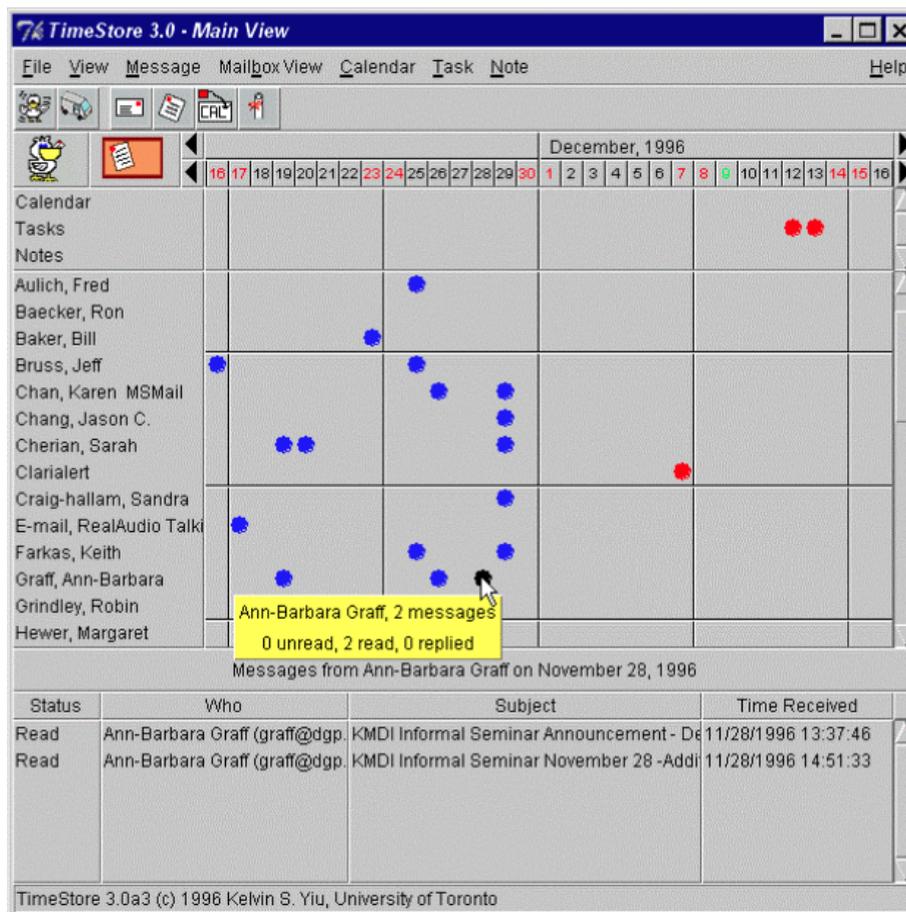


Figura 3.3: Imagem do TimeStore retirada da publicação de Kelvin Yiu et al. [YBSL97].

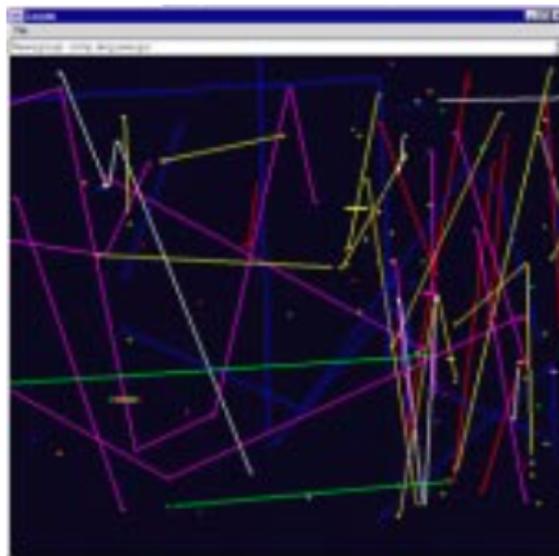


Figura 3.4: Imagem da interface do Loom retirada da publicação de Judith Donath, Karrie Karahalios e Fernanda Viegas [YBSL97].

E-mail

Steven Rohall et al. [RGMK01, Roh02] abordaram a visualização de uma outra forma. Eles pretendiam criar uma visualização que destacasse a relação entre as mensagens e entre as pessoas que trocam essas mensagens. Dessa forma eles implementaram um novo cliente de e-mail, que é mostrado na Figura 3.5, que combinava uma tradicional lista de mensagens de correio eletrônico com uma árvore temporal de mensagens. É ainda de referir que os mesmos autores faziam parte de um grupo de investigação da IBM (*International Business Machines*) que fez diversas pesquisas e protótipos de clientes de e-mail num projeto conhecido por *Reinventing Email (Re-Mail)* [RGM⁺04, Moo03]. Este projeto além de conter a já referida lista de mensagens complementada com uma árvore temporal, para identificar as interações entre as mensagens nas *threads*, implementava ainda uma visualização que permitia ver a distribuição dos resumos de cada mensagem pelos seus dias de recepção, e ainda um sistema de *chat*.

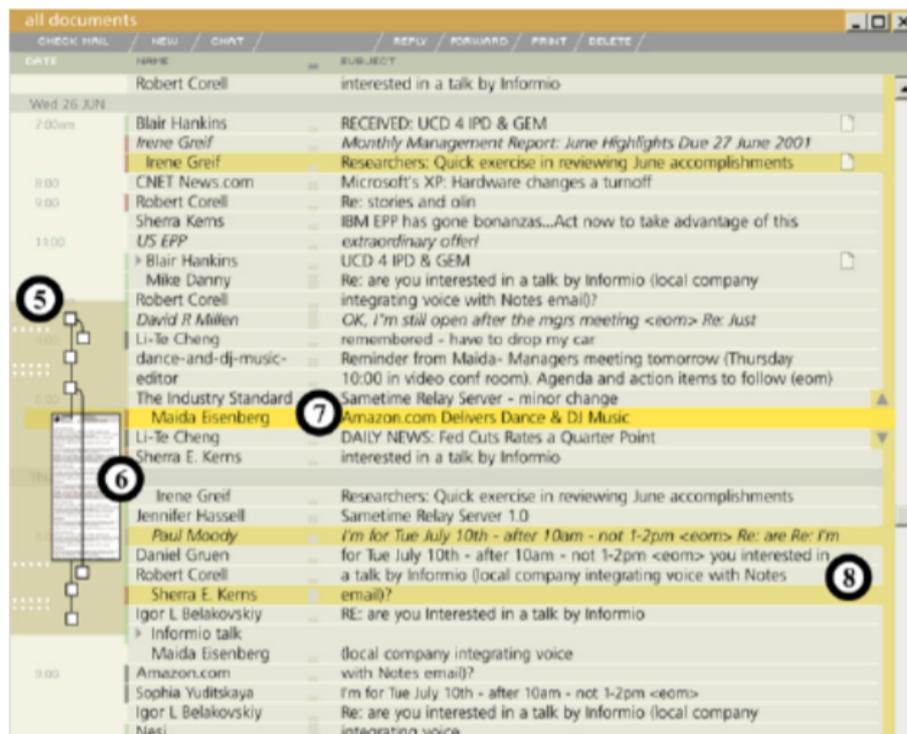


Figura 3.5: Imagem da interface para o e-mail proposta por Steven Rohall et al. [RGMK01, Roh02].

Uma outra publicação da autoria de Olle Balter e Candace Sidner [BS02] sugeriu uma forma de visualização menos disruptiva que as anteriores mas com boa aceitabilidade por parte dos utilizadores. Eles sugeriam uma interface que priorizava as mensagens de correio eletrônico em diversas categorias segundo regras predefinidas. As categorias agrupavam mensagens que continham referências temporais ao dia em curso, a remetentes importantes definidos pelos utilizadores, a mensagens cujo o único destinatário era o próprio utilizador e ainda uma categoria para mensagens que continham vários destinatários. Esta interface é mostrada na Figura 3.6.

E-mail

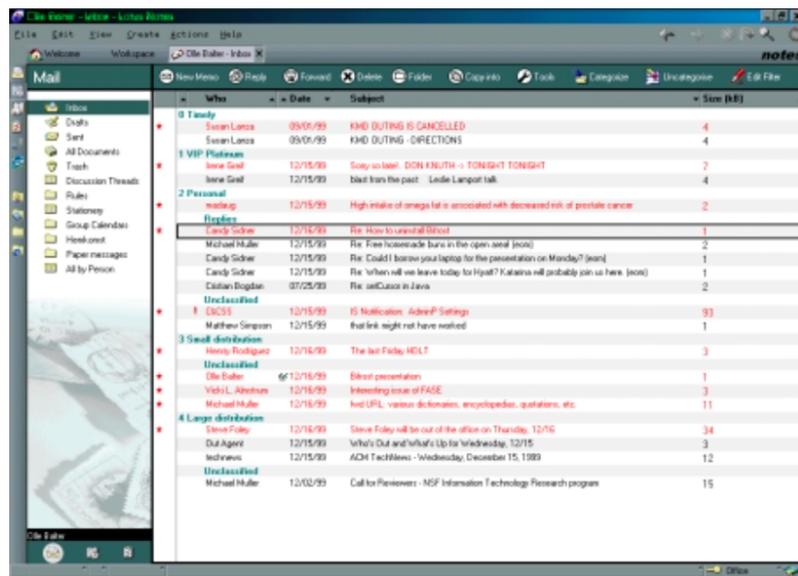


Figura 3.6: Imagem da interface para o e-mail proposta por Olle Balter e Candace Sidner [BS02].

Jacek Gwizdka [Gwi02a, Gwi02b] baseado no trabalho de Kelvin Yiu et al. [YBSL97], o TimeStore, sugeriu também uma visualização para um cliente de e-mail, que tinha o objetivo de melhorar a gestão de tarefas. A visualização proposta pelo autor, e que é mostrada na Figura 3.7, representava as tarefas como pequenos ícones numa grelha de duas dimensões que continha o tempo no eixo dos x e outras informações como o assunto e o remetente no eixo dos y. Uma particularidade desta visualização é de apenas serem mostradas as tarefas pendentes, que tenham qualquer tipo de referência futura.

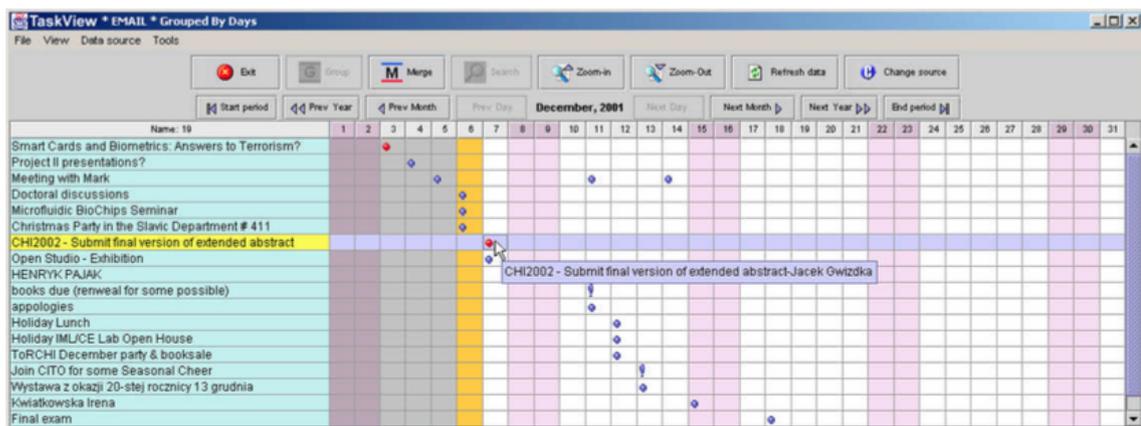


Figura 3.7: Imagem da interface para o e-mail proposta por Jacek Gwizdka [Gwi02a, Gwi02b].

Continuando a nossa análise ao que já foi realizado na área, a publicação de Sandra Sudarsky e Rune Hjelmsvold [SH02] sugere uma abordagem visual para o e-mail capaz de conferir uma classificação automática para as mensagens com base nos domínios dos endereços de e-mail. A interface conta ainda com uma vista temporal que distribui as mensagens numa grelha de duas

E-mail

dimensões, que contém o tempo no eixo dos x e o remetente no eixo dos y. Um problema apontado a este tipo de visualização prende-se com o facto de ser comum a receção de várias mensagens de um mesmo domínio de correio eletrónico o que reduz em muito a eficiência desta visualização.

Victoria Bellotti et al. [BDHS03] dão conta na sua publicação de um *plugin* para o cliente de e-mail Outlook, de seu nome Taskmaster, que agrupa mensagens que tenham tarefas semelhantes associadas, assim como anexos. Os autores apelidam estas mensagens agrupadas de *thrasks*, e estas são apresentadas no painel principal da sua interface. Depois o painel do meio apresenta toda a informação associada à *thrask*, seja ela mensagens, anexos ou até mesmo *links*. O último painel apresenta, por fim, o conteúdo de uma dada mensagem. A interface permite ainda que sejam adicionadas etiquetas, lembretes, prazos de entrega ou outro tipo de informações às *thrasks*. Uma imagem desta interface pode ser consultada na Figura 3.8.

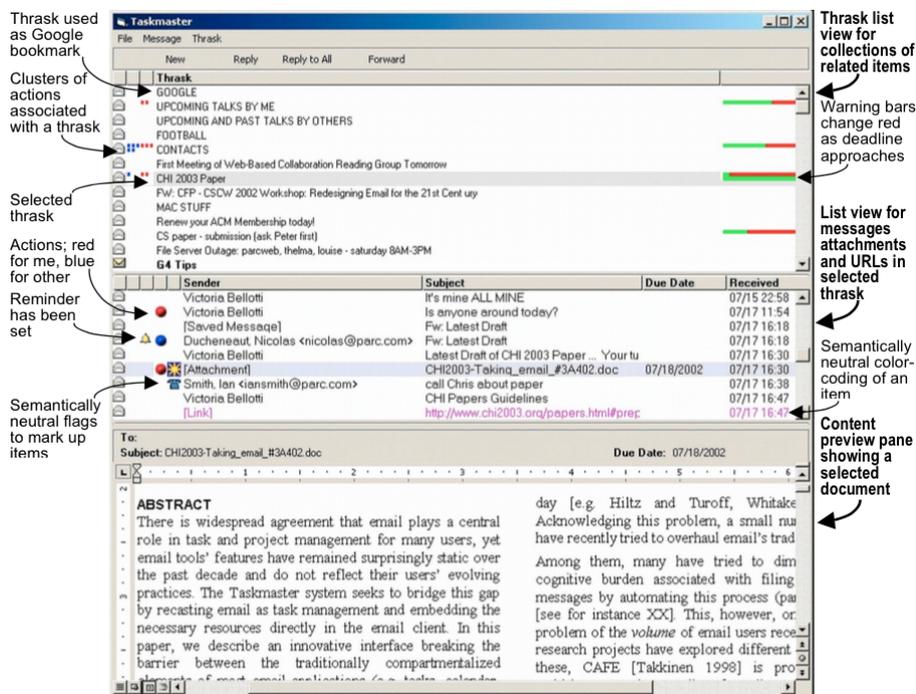


Figura 3.8: Imagem da interface para o e-mail proposta por Victoria Bellotti et al. [BDHS03].

Ainda no mesmo ano, aparece uma nova abordagem à visualização no e-mail, também ela associada ao projeto *Reinventing Email (ReMail)* [RGM⁺04, Moo03]. A publicação é da autoria de Bernard Kerr [Ker03] e denomina-se por *ThreadArcs*. O *ThreadArcs* é uma técnica de visualização de *threads* que torna visual a cronologia de uma *thread* através da criação de nós de mensagem num esquema bem definido, e que ganha contexto através dos arcos de relação que ligam os nós. Esta visualização permite facilmente adquirir um conjunto de informação fundamental num contexto de uma *thread* como por exemplo as mensagens ainda não lidas, ou ainda as mensagens enviadas por um dado remetente. A técnica de visualização pode ser vista na Figura 3.9.

E-mail

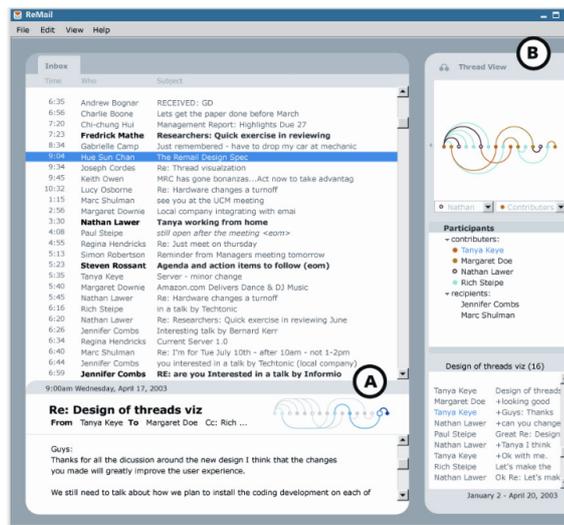


Figura 3.9: Imagem da técnica de visualização *ThreadArcs* proposta por Bernard Kerr [Ker03].

Gina Venolia e Carman Neustaedter [VN03] publicaram também eles uma abordagem à visualização do correio eletrónico diferente. Os autores apresentaram um modelo misto que suporta dois tipos de conversação: um modelo sequencial que cria uma lista cronológica de mensagens agrupadas por conversa, e um modelo em árvore que especifica a sequência de respostas numa *thread*. Este modelo permite que ao mesmo tempo que se observa a conversa, pela ordem natural na qual ela ocorreu, ainda se consiga ter um contexto de onde se insere a *thread* na nossa caixa de correio. A técnica de visualização pode ser vista na Figura 3.10.

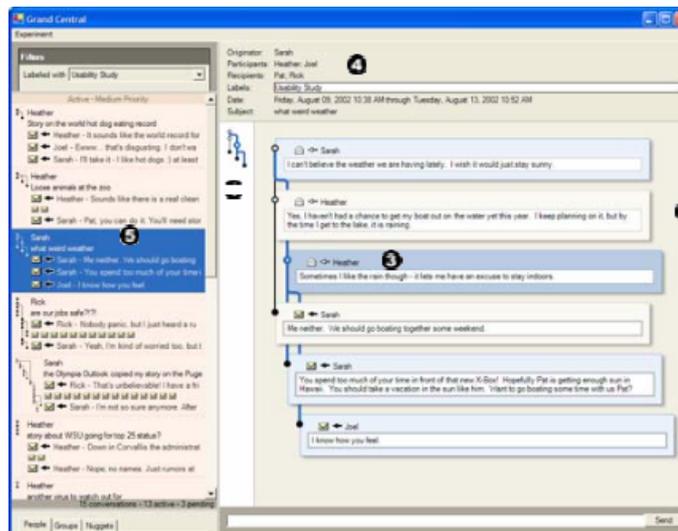


Figura 3.10: Imagem da visualização proposta por Gina Venolia e Carman Neustaedter [VN03].

A publicação de Mirko Mandic e Android Kerne [MK04] sugere uma abordagem baseada em intimidade. Na sua abordagem, os autores consideram dois tipos de intimidade: a intimidade

E-mail

do contacto e a intimidade da mensagem. A sua visualização agrupa de forma cronológica as mensagens marcando depois as mesmas com cores e símbolos diferentes dependendo do seu grau de intimidade para com o utilizador. Os autores afirmam ainda que este método de visualização aplicado ao correio eletrónico permite a deteção de ritmos e padrões de comunicação, que de outra forma permaneceriam ocultos. As funcionalidades de pré-visualização e de filtragem de mensagens estão também contempladas nesta interface. A referida interface de visualização é apresentada na [Figura 3.11](#).

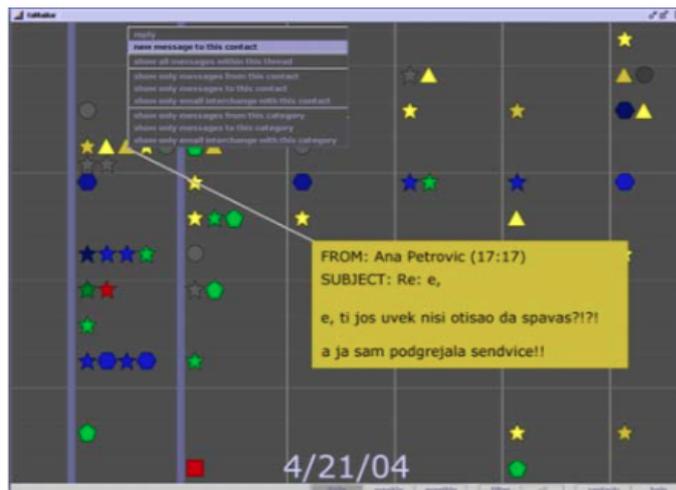


Figura 3.11: Imagem da visualização proposta por Mirko Mandic e Android Kerne [\[MK04\]](#).

Posteriormente Victoria Bellotti e Jim Thornton [\[BT06, BTC+07\]](#) divulgaram um trabalho onde dão a conhecer um protótipo que desenvolveram chamado *TaskVista*, que implementa uma lista de tarefas num cliente de e-mail. O protótipo dos autores permite aos utilizadores criarem tarefas arrastando ficheiros ou mensagens. O *TaskVista* incorpora também um gestor de atividades conhecido por ACTA (*Activity-Centered Task Assistant*) que permite aos utilizadores agruparem várias tarefas numa atividade. O sistema está representado na [Figura 3.12](#).

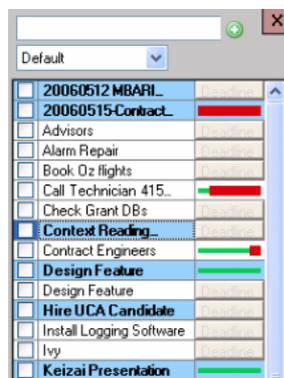


Figura 3.12: Imagem da visualização proposta por Victoria Bellotti e Jim Thornton [\[BT06, BTC+07\]](#).

Por fim, a terminar esta análise aos métodos de visualização já aplicados ao e-mail, analisa-se o trabalho de Romain Vuillemot, Jean Petit e Mohand Hacid [VPH10], que dá pelo nome de *SHIFT-BOX*. Neste trabalho os autores exploraram a ideia de dar aos utilizadores a possibilidade de eles poderem saltar entre os tempos de receção das mensagens, assim como fazer pausa ou repetir um determinado fluxo de mensagens numa janela temporal. O resultado do trabalho assemelha-se claramente a um reprodutor de ficheiros de áudio aplicado a um cliente de e-mail, e com ele os autores quiserem testar se o mesmo poderia ou não ajudar a resolver os problemas de sobrecarga sobre os utilizadores, verificado no e-mail. A Figura 3.13 mostra claramente todas as funcionalidades do sistema aqui referidas.

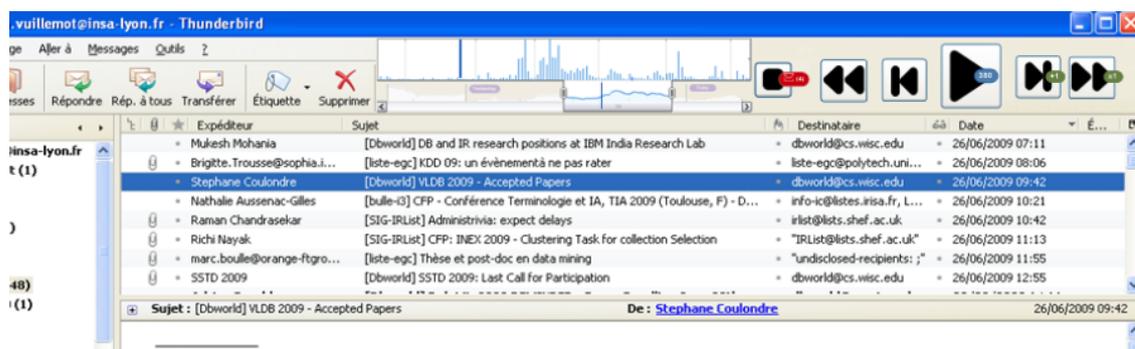


Figura 3.13: Imagem da visualização proposta por Romain Vuillemot, Jean Petit e Mohand Hacid [VPH10].

3.2 Aprendizagem e Extração de Conhecimento

O e-mail assumiu um papel importantíssimo na sociedade e tem sofrido de problemas já há vários anos, que se têm mantido por resolver [WBM05, WS96]. A aprendizagem e extração de conhecimento tem-se dedicado nos últimos anos ao estudo de diversas áreas como o jornalismo, biomedicina e é claro também ao e-mail. Algumas publicações têm sugerido que o uso de algoritmos de aprendizagem e de extração de conhecimento aplicados ao e-mail podem resolver alguns dos problemas identificados.

Por isso, será agora realizada uma revisão ao que de relevante foi publicado envolvendo o e-mail e a Aprendizagem e Extração de Conhecimento, sendo que se dividiu a análise em dois temas: Classificação e Sumarização.

3.2.1 Classificação

No trabalho já publicado sobre o e-mail, a classificação foi estudada e aplicada desde muito cedo. Em trabalhos publicados, e já analisados ao longo deste capítulo anteriormente, acredita-se que uma classificação eficaz possa ajudar a resolver o problema do e-mail. Nesse sentido, vamos passar agora a analisar as publicações relacionadas com algoritmos e técnicas de classificação usadas ou passíveis de serem aplicadas ao e-mail.

O primeiro trabalho a analisar data de 1996 e é de William Cohen [Coh96]. Neste trabalho o autor compara dois algoritmos de classificação de texto TF-IDF (*Term Frequency–Inverse Document Frequency*) [Sal91] e o RIPPER (*Repeated Incremental Pruning to Produce Error Reduction*) [Coh95]. O autor refere que os resultados das suas experiências são bastante encorajadores, uma vez que a curva de aprendizagem destes algoritmos é muito pequena. Consequentemente, segundo o mesmo, estes algoritmos são totalmente passíveis de ser aplicados ao problema do e-mail. É ainda deixada a ideia de que um sistema que consiga combinar uma aprendizagem orientada ao utilizador construída com recurso a regras que encontrem palavras-chave pode ser uma arquitetura viável para um sistema de classificação e filtragem de e-mails.

Richard Segal e Jeffrey Kephart [SK99] também publicaram um estudo no qual afirmam que a maior parte dos clientes de e-mail deixam os utilizadores organizarem as mensagens em pastas, o que tende a ser uma tarefa muito dispendiosa. Eles afirmam também que o processo de ter de ler a mensagem e perceber em qual pasta a organizar desencoraja por completo os utilizadores a organizarem a sua caixa de correio eletrónico. Para resolver o referido problema eles propõem uma solução chamada MailCat que, recorrendo a um classificador adaptativo, é capaz de prever as três pastas às quais uma mensagem é mais provável pertencer. Depois deste processo, o MailCat oferece botões de acesso rápido para facilmente o utilizador escolher uma das três pastas sugeridas. Os resultados apontam para que o MailCat tenha uma eficácia de 80% a 90%.

Elisabeth Crawford, Judy Kay e Eric McCreath [CKM02] apresentaram uma abordagem mais comum. No seu trabalho eles falam da interface *i-ems* que oferece uma maneira de categorizar e ler e-mails. Os autores seguiram uma abordagem convencional de um cliente de e-mail, com uma painel lateral esquerdo que lista as pastas do e-mail, um painel central superior que lista as mensagens dentro de uma pasta selecionada, e um painel central inferior que mostra o conteúdo de uma dada mensagem. A novidade trazida por esta publicação ocorre na vista que lista as mensagens dentro da caixa de entrada. As mensagens dentro da caixa de entrada aparecem ordenadas segundo a categorização determinada pela solução. Segundo os autores, no caso de os algoritmos de classificação falharem, o trabalho que o utilizador enfrenta de ter de categorizar a mensagem, é o mesmo que teria se não usasse a ferramenta. Portanto, usá-la é benéfico.

Outros autores foram um pouco mais longe e sugeriram uma classificação substancialmente diferente. Mark Dredze e Tessa Lau [DLK06] sugeriram um sistema capaz de classificar mensagens de e-mail em atividades. Os autores criaram uma extensão para o cliente de e-mail ThunderBird⁶ que cria um painel lateral que resume as atividades tratadas numa dada mensagem. Esta extensão recorre ao uso dos algoritmos SimSubset e SimOverlap, também criados pelos autores, que conseguem uma classificação das mensagens em atividades com uma eficácia muito alta de 80%.

Uma outra publicação muito interessante tem a autoria de Douglas Aberdeen, Ondrej Pacovsky e Andrew Slater [APS10] investigadores da Google Inc. Os investigadores sugerem um modelo misto que combine um modelo estatístico global, criado com base na gigante quantidade de dados a que a Google tem acesso, complementado por um modelo estatístico do utilizador, atualizado o mais frequentemente possível. Nos testes realizados à sua abordagem, os autores afirmam que

⁶<https://www.mozilla.org/pt-PT/thunderbird/>

o seu sistema oferece uma eficácia de cerca de 80% e que, em testes realizados a cerca de 2000 utilizadores da Google, permite uma redução de 6% no tempo gasto a ler o e-mail de forma geral, e menos 13% de tempo gasto a ler e-mail não importante.

Por fim, Anton Borg e Niklas Lavesson [BL12] abordaram a classificação de e-mail, para um fim um pouco diferente, o de filtragem de lixo eletrónico. Os autores sugerem um sistema capaz de filtrar o lixo eletrónico recorrendo à informação presente nas redes sociais. No entanto, os seus resultados ditam que, esta solução se mostra bastante mais eficaz quando aplicada em forma de complemento às soluções já utilizadas, ao invés de ser usada individualmente.

3.2.2 Sumarização

A sumarização de textos é uma técnica comum no âmbito da aprendizagem e da extração de conhecimento. No caso do e-mail, com o crescendo de mensagens recebidas, começaram a aplicar-se este tipo de algoritmos com o objetivo de dar ao utilizador a perceção do conteúdo de dada mensagem, sem no entanto este a ler por completo. No entanto, o correio eletrónico apresentou diversas dificuldades ao uso destas técnicas uma vez que as mensagens do mesmo são compostas por diversos autores, *threads* e muitas vezes por abreviaturas ou estilos de escrita menos comuns.

O primeiro trabalho a ser analisado é o de Smaranda Muresan, Evelyne Tzoukermann e Judith Klavans [MTK01] e os autores apresentam uma abordagem na qual combinam técnicas linguísticas e de aprendizagem para extrair não frases de qualidade elevada. As não frases extraídas têm como objetivo fornecer um resumo confiável das mensagens de e-mail a que dizem respeito. Os autores acentuam ainda as dificuldades de aplicar estes algoritmos a textos provenientes de mensagens de e-mail, uma vez que estes por norma não são estruturados e nem sempre estão sintaticamente bem construídos. As conclusões do trabalho são essencialmente três: os modificadores das não frases são semanticamente importantes, a filtragem linguística melhora a performance dos algoritmos de aprendizagem e, por fim, a combinação de classificadores melhora a exatidão dos resultados.

Uma outra abordagem à temática da sumarização foi abordada por Paula Newman e John Blitzer [New02, NB03] nas suas publicações, a primeira tendo como autoria Paula Newman e a segunda sendo já uma publicação conjunta. Na sua solução os autores propõem um algoritmo capaz de resumir *threads* de discussão arquivadas, agrupando as mensagens em grupos de assunto semelhante, e posteriormente extraíndo pequenos e longos resumos para cada grupo. Nas suas conclusões Newman e Blitzer referem que os resultados de teste da sua solução foram encorajadores mas que ainda era preciso bastante trabalho para melhorar os métodos e avaliar a sua utilidade para os utilizadores finais.

Owen Rambow et al. [RSCL04] publicaram também um trabalho dentro desta temática. Na sua publicação os autores referem que na abordagem escolhida usaram um paradigma já usado para outros tipos de sumarização, o da extração de frases. Este paradigma baseia-se na ideia de as frases importantes serem extraídas das *threads* e serem interligadas entre elas para formar um

resumo. Eles também preveem que, dadas as características do e-mail, existem pequenas particularidades que podem ajudar a identificar as frases a serem extraídas. Por fim, sublinham ainda que manter a coerência de interação dentro da *thread* é fundamental para obter bons resultados.

Continuando a nossa análise encontra-se um outro trabalho de Simon Oliver et al. [CORGC04] que aborda o problema da sumarização de uma maneira bastante diferente. No trabalho citado é apresentado um sistema que identifica automaticamente tarefas contidas numa determinada mensagem de e-mail. O sistema começa por analisar superficialmente a mensagem de correio eletrónico para distinguir o cabeçalho da mensagem. Depois cada frase é analisada por um classificador que identifica se a frase pode ou não ser classificada como uma tarefa. Posteriormente, o sistema referido reformula as frases identificadas como tarefas e apresenta-as ao utilizador numa interface conveniente, a partir da qual podem facilmente adicioná-las a uma lista de tarefas.

Angelo Dalli, Yunqing Xia e Yorick Wilks [DXW04] contribuíram também com o seu trabalho para este tema, sendo este, no entanto, substancialmente diferente dos anteriores. Eles apresentaram um serviço de sumarização, compacto e otimizado, para ser aplicado num sistema baseado em voz. O seu sistema tinha a particularidade de ser rápido e adaptativo o suficiente para poder ser integrado num assistente pessoal virtual para telemóveis, que tinha como objetivo a gestão da informação pessoal e profissional dos seus utilizadores. O objetivo principal do sistema era o de conseguir aumentar a produtividade dos utilizadores, priorizando e apresentando em dado momento a informação mais pertinente, adaptando-se também de forma individual às preferências de cada um.

Stephen Wan e Kathy McKeown [WM04] por sua vez apresentaram um sistema capaz de resumir conversas no correio eletrónico, que dava especialmente enfoque aquelas que não estavam ainda arquivadas, ou seja, que estavam a decorrer. O objetivo principal dos autores era com o seu trabalho oferecer uma maneira de auxiliar as decisões que tinham de ser tomadas em determinada conversação. De referir também que, os resumos criados ofereciam uma pré-visualização do estado atual da conversa ao mesmo tempo que auxiliavam o utilizador a responder rapidamente, contribuindo dessa forma para ajudar a resolver um dos problemas do e-mail. A técnica utilizada recorria ao uso da estrutura da *thread* assim como à técnica do vetor de palavras para determinar qual a frase que deveria ser extraída.

Por sua vez Giuseppe Carenini, Raymond Ng e Xiaodong Zhou [CNZ07, CNZ08] no seu trabalho referem que o elevado número de mensagens de e-mail recebidas, principalmente em dispositivos móveis, se tornou um verdadeiro problema, e que uma forma de o resolver poderá ser através da sumarização de mensagens. A sua solução passa pela utilização de um grafo de citações que tem como objetivo capturar da melhor forma as conversas que vão decorrendo no e-mail. De forma complementar, usam também palavras-chave para tentaram medir a importância de cada frase num resumo. A grande particularidade desta solução é que permite ao utilizador definir o tamanho que pretende que o resumo tenha.

De seguida, analisaremos o trabalho proposto por Lokesh Shrestha, Kathleen McKeown e Owen Rambow [SMR07]. No seu trabalho os autores referem que a extração de frases demonstrou ser um método funcional no resumo de documentos de dado tipo, mas que poderia não resultar no

contexto textual do e-mail. Dessa forma propõem-se a fazê-lo através de uma detecção automática de pares questão-resposta. Os autores fizeram ainda diversas comparações do seu método de questão-resposta e demonstraram a sua qualidade face a outros. Por fim no seu trabalho, descrevem ainda a interface que desenvolveram para o cliente de e-mail *Microsoft Outlook* que permite aos utilizadores testarem esta abordagem de sumarização.

Taiwo Ayodele, Shikun Zhou e Rinat Khusainov [AZK09] propuseram um sistema capaz de agrupar e resumir as mensagens de correio eletrónico. O seu trabalho refere que o sistema utilizava o assunto e o conteúdo das mensagens, para classificar os e-mails com base nas atividades do utilizador e, de forma automática, gerar resumos para cada mensagem que chega à caixa de correio. Segundo os autores, a sua solução ajuda a resolver claramente os problemas de sobrecarga, congestão e as dificuldades de priorização verificadas no e-mail, permitindo também um processamento mais rápido das diversas mensagens, assim como uma pesquisa melhorada àquelas que já estão arquivadas. Por fim, referem ainda que todo o sistema assenta num algoritmo bastante simples mas que, ao mesmo tempo, é bastante complexo de ser implementado.

Por fim, e a terminar esta análise às soluções e aos trabalhos publicados sobre técnicas de sumarização analisamos brevemente outros dois trabalhos. O primeiro foi publicado por Jan Ulrich et al. [UCMN09] e descreve um sistema de sumarização com base numa aprendizagem usando modelos de regressão. De acordo com os autores, o modelo de regressão é capaz de retirar vantagem das múltiplas anotações que podem ser efetuadas com objetivo de treino, em contraste ao que habitualmente é usado: classificadores binários. O segundo foi publicado por Shixia Liu et al. [LZP⁺09] e descreve um sistema, de seu nome TIARA, a partir do qual é possível resumir mensagens de e-mail de uma forma visual. O sistema apresenta diversas palavras-chave num gráfico visual, usando diferentes cores e tamanhos para destacar as diferentes importâncias de cada palavra. O utilizador pode ainda consultar excertos de texto, ou até mesmo o texto original que deu origem a determinada parte do resumo.

3.3 Trabalhos Industriais

Depois de analisadas as publicações académicas, não seria possível fazer um estudo do estado da arte completo, sem também analisar os produtos comerciais no mercado. Isto porque, se a investigação é uma parte essencial para a resolução de um dado problema, aquilo que já está de facto a ser usado e disponível para as pessoas é também crucial, não só para compreender o problema, mas também para perceber outros modos de o abordar, caso ele ainda não esteja completamente resolvido. Ao longo da pesquisa efetuada foram encontrados bastantes produtos, sendo que a grande maioria será agora analisada. A análise não pretende ser exaustiva mas sim abordar aquilo que cada um tem de mais importante para oferecer. No fim da análise individual será apresentada uma comparação entre cada um deles resumida sob a forma de uma tabela.

3.3.1 SaneBox

O SaneBox⁷ é um serviço web pago que monitoriza caixas de e-mail IMAP, e que executa algumas ações inteligentes acerca das mensagens que vão ou não para a caixa de entrada. Se o serviço perceber que dada mensagem de e-mail não é necessária redireciona-a para outras pastas presentes no e-mail para que o utilizador, caso pretenda, possa verificar ou ler mais tarde. Este comportamento permite, ao mesmo tempo, que a caixa de correio eletrónico se mantenha mais vazia e que o utilizador não perca nenhuma mensagem. As diversas pastas já referidas, e utilizadas pelo SaneBox, são por defeito a SaneLater que é responsável por mensagens genéricas que podem ser vistas mais tarde, a SaneBulk para mensagens que são enviadas por serviços de envio em massa, a SaneNews para mensagens que dizem respeito a newsletters e ainda a SaneBlackHole que faz com que remetentes que contenham mensagens dentro desta pasta vejam as suas novas mensagens irem diretamente para o lixo. Adicionalmente importa referir que o SaneBox vai tentando atualizar as suas regras de filtragem de acordo com as ações que o utilizador vai fazendo no e-mail. A grande desvantagem do SaneBox é o facto de o mesmo ser apenas um serviço e não oferecer nenhum tipo de cliente de e-mail associado. Por fim, este é um serviço que tem um custo uma vez que apresenta planos de pagamento periódicos. A sua interface pode ser consultada na Figura 3.14.

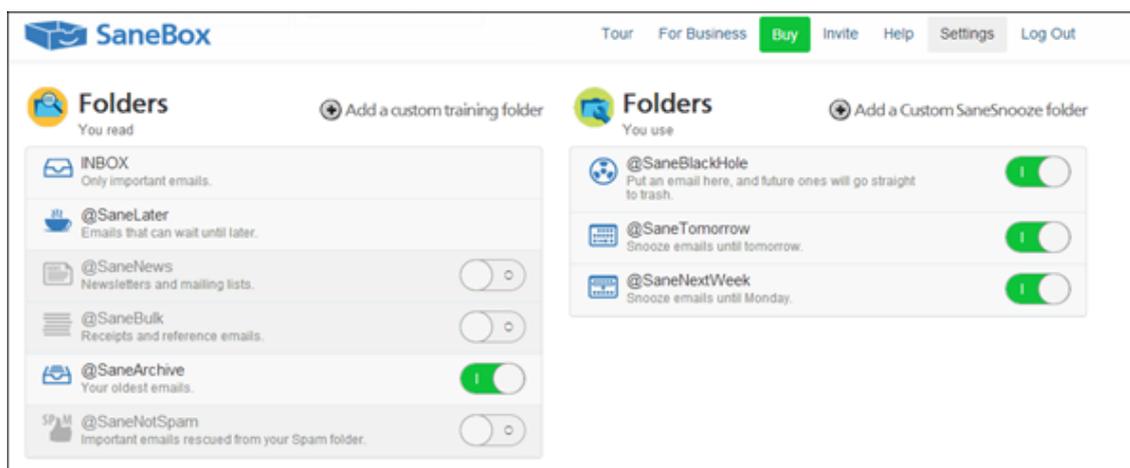


Figura 3.14: Imagem da interface do SaneBox.

3.3.2 ActiveInbox

O ActiveInbox⁸ é um plugin para navegadores de internet desenhado para o Gmail, que permite transformar o e-mail numa ferramenta de gestão de tarefas. Este serviço permite adicionar tarefas às mensagens de e-mail, assim como organizá-las numa lista de tarefas que segue um determinado escalonamento definido pelo próprio utilizador, baseado nos prazos de entrega e na disponibilidade do mesmo. Permite ainda o agendamento de mensagens para serem vistas mais tarde, permite

⁷<https://www.sanebox.com>

⁸<http://www.activeinboxhq.com>

E-mail

marcar mensagens que têm de ser respondidas ou mensagens das quais esperamos uma resposta e ainda a criação de lembretes. A grande desvantagem deste produto é o facto de o mesmo só suportar o provedor Gmail e ainda o facto de resumir o uso do e-mail para uma vertente única de gestão de tarefas. É um serviço pago com diferentes planos que pretendem abranger um conjunto vasto de utilizadores. A sua interface pode ser consultada na [Figura 3.15](#).

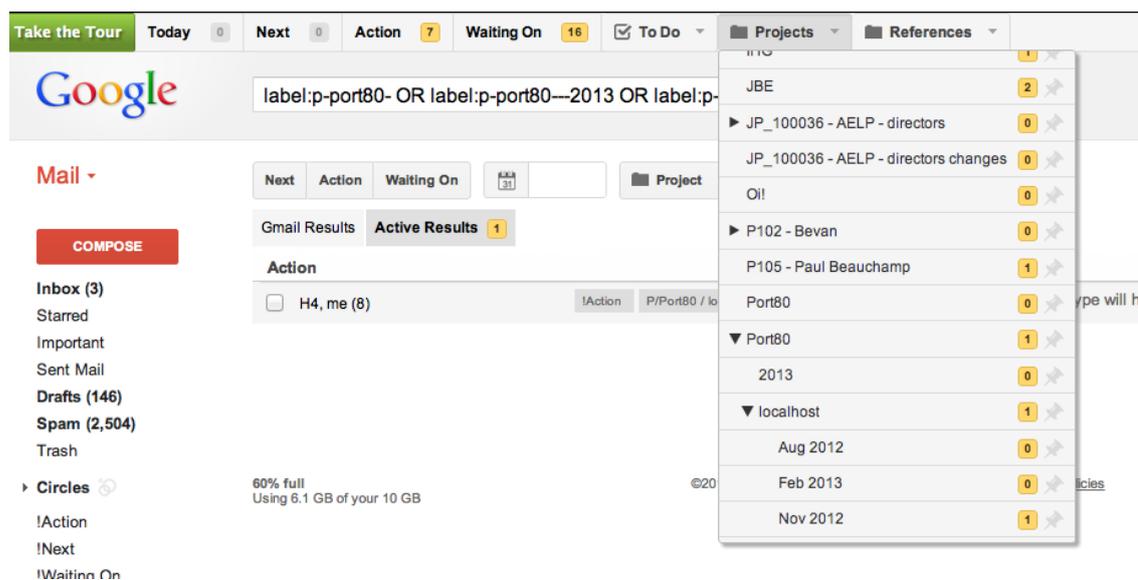


Figura 3.15: Imagem da interface do ActiveInbox.

3.3.3 Inboxcube

O Inboxcube⁹ é um cliente de e-mail sob a forma de uma aplicação móvel para iOS¹⁰, que chegou com a promessa de mudar bastante a maneira como o utilizador visualiza e interage com o e-mail. Esta aplicação tornou-se bastante conhecida após se ter tornado vencedora de um *Appy Award*¹¹ na categoria de *Best Communication App*. Foi pensada para funcionar com provedores de e-mail IMAP e a sua visualização assenta numa ideia de cubo. A aplicação organiza determinada informação num lugar único e organizado, como por exemplo fotos, ficheiros PDF (*Portable Document Format*), *links* ou ainda contactos importantes e pastas. A grande desvantagem deste cliente de e-mail prende-se com o facto de o mesmo, atualmente, só estar disponível em dispositivos iOS. A aplicação é gratuita e a sua interface pode ser consultada na [Figura 3.16](#).

⁹<http://www.inboxcube.com>

¹⁰<https://www.apple.com/ios/>

¹¹<http://www.mediapost.com/appyawards/>

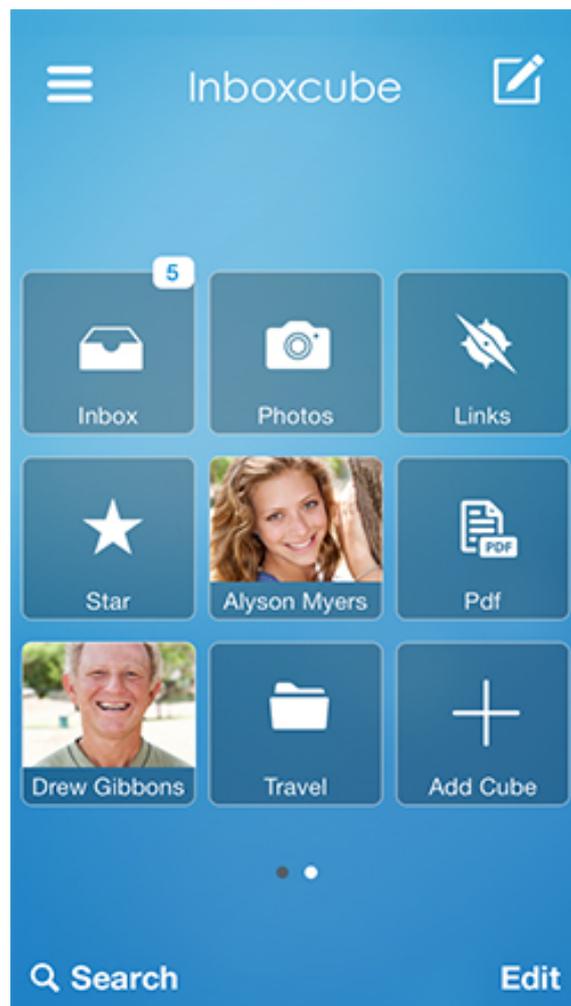


Figura 3.16: Imagem da interface do Inboxcube.

3.3.4 Inbox by Gmail

O Inbox by Gmail¹² é um cliente de e-mail com versão web e versão para dispositivos móveis. O grande objetivo do Inbox by Gmail é ser capaz de mostrar a informação importante para o utilizador, de uma forma contextual, ao invés de a apresentar à medida que ela chega. As principais características do cliente prendem-se com o mesmo agrupar de forma conjunta as mensagens semelhantes, mostrar sem abrir a mensagem a informação mais importante, dar a possibilidade de adicionar lembretes e ainda o facto de permitir que mensagens e lembretes voltem a ser entregues mais tarde. Contudo, uma das grandes desvantagens deste cliente de e-mail é o facto de o mesmo apenas funcionar com o serviço de e-mail Gmail. O Inbox by Gmail é grátis e a sua interface pode ser consultada na [Figura 3.17](#).

¹²<http://www.google.com/inbox/>

E-mail

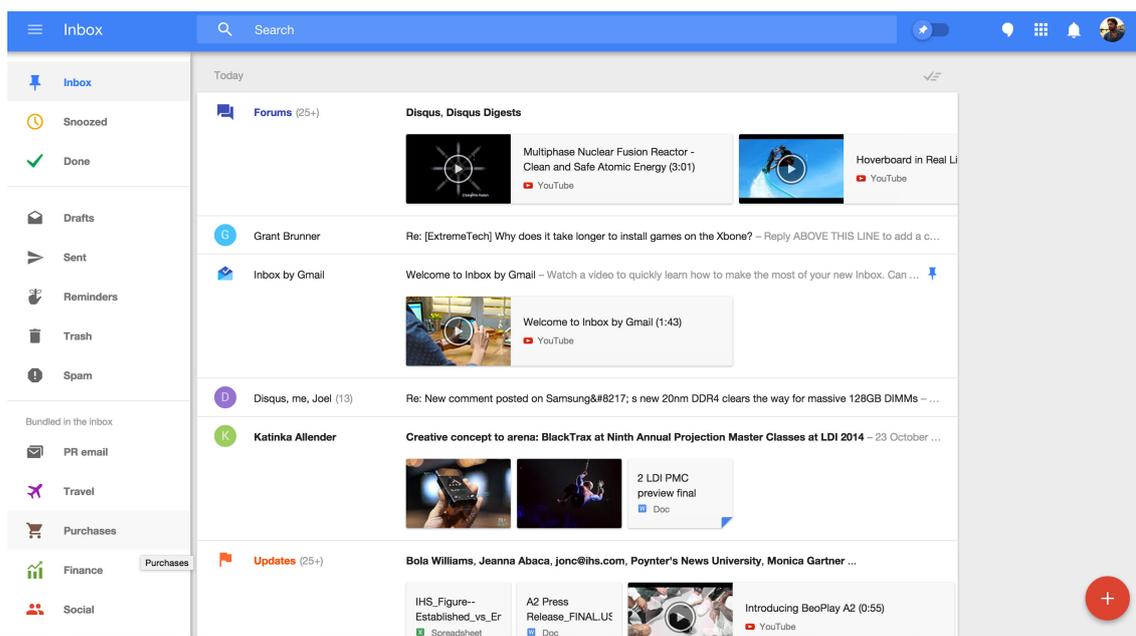


Figura 3.17: Imagem da interface do Inbox by Gmail.

3.3.5 Alto by AOL

O Alto by AOL¹³ (*America Online*) é um cliente de e-mail grátis, exclusivamente com versão web, que funciona com alguns dos mais conhecidos provedores de e-mail IMAP. O Alto introduziu no mercado uma diferente visualização no e-mail, uma vez que ele coleciona as mensagens em pilhas virtuais. As pilhas criadas por defeito são as das promoções diárias, notificações sociais, fotos e anexos, sendo possível a criação de pilhas adicionais. Uma funcionalidade adicional do Alto é a possibilidade de o mesmo arquivar temporariamente uma mensagem voltando-a a entregar numa altura mais conveniente, definida pelo utilizador. Uma das grandes desvantagens do alto prende-se com o facto de não funcionar com todos os provedores de e-mail IMAP e de permitir apenas uma definição muito básica nas regras que ditam a criação das pilhas virtuais. A interface do Alto by AOL pode ser consultada na [Figura 3.18](#).

¹³<http://altomail.com>

E-mail

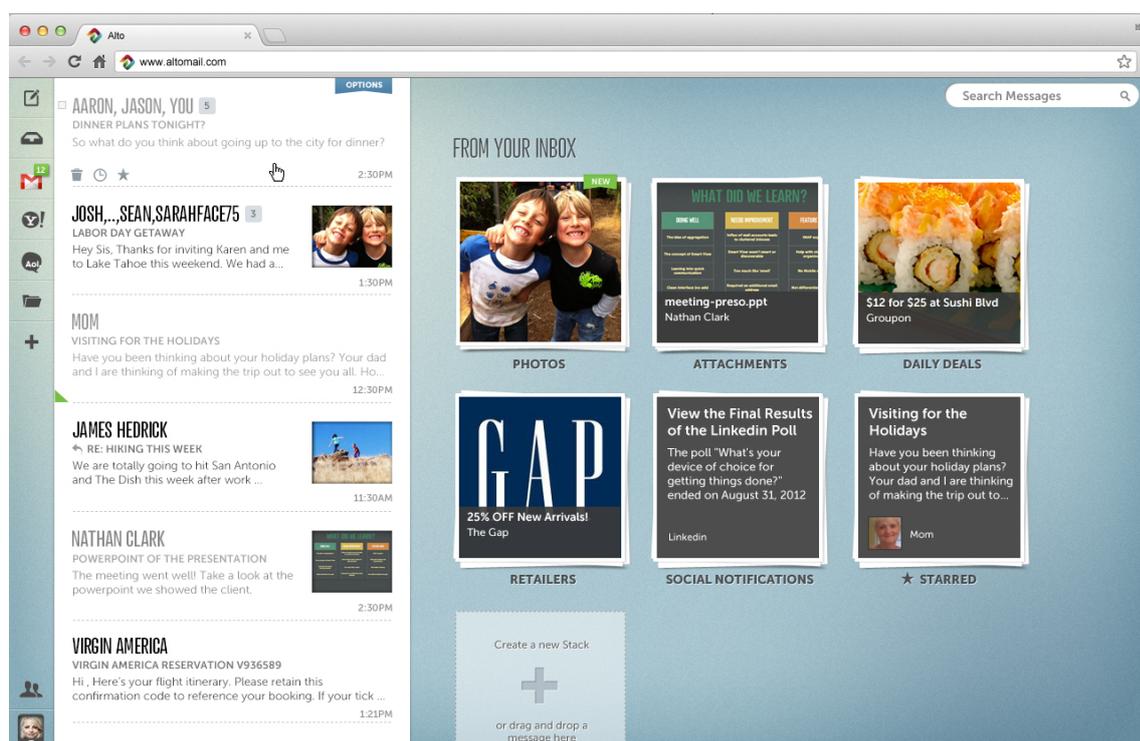


Figura 3.18: Imagem da interface do Alto by AOL.

3.3.6 Mailbox

O Mailbox¹⁴ é um cliente de e-mail sob a forma de aplicação móvel e desktop (atualmente apenas disponível para OSX¹⁵) devido pela Dropbox¹⁶. O referido cliente de e-mail tem como objetivo limpar por completo a caixa de entrada do utilizador, ou pelo menos ajudá-lo a geri-la melhor. Aquando do seu lançamento introduziu diversas novidades, especialmente na forma como os utilizadores interagem com o cliente de e-mail recorrendo a diversos atalhos para facilmente ser possível tratar de uma mensagem. Disponibiliza funcionalidades importantes como a de adicionar uma mensagem de e-mail a uma determinada lista ou ainda a de reagendar a entrega de uma mensagem. A grande desvantagem deste cliente de e-mail prende-se com o facto de não disponibilizar nenhum cliente web e de oferecer suporte para poucos provedores de e-mail, pelo menos atualmente. O Mailbox é uma aplicação gratuita e a sua interface pode ser consultada na Figura 3.19.

¹⁴<http://www.mailboxapp.com>

¹⁵<https://www.apple.com/osx/>

¹⁶<https://www.dropbox.com/>

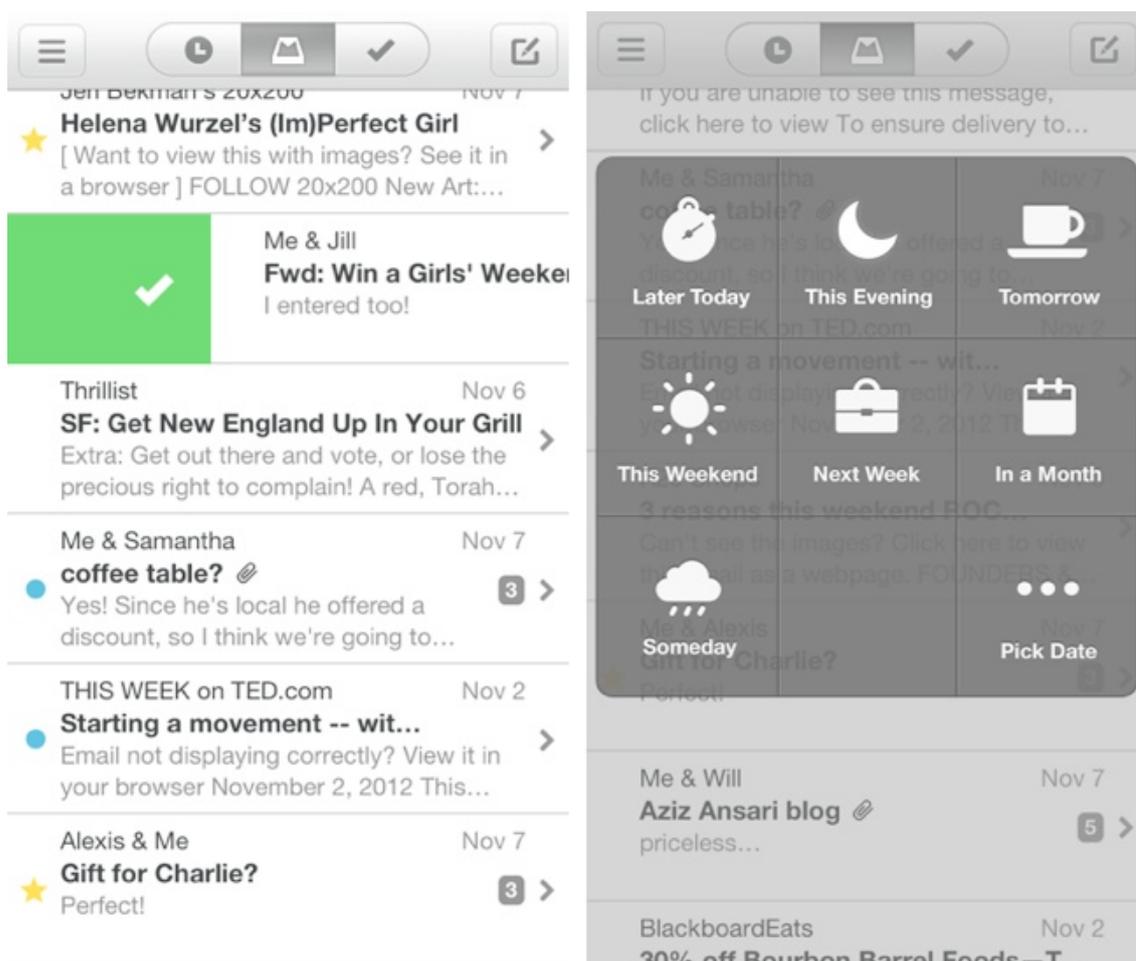


Figura 3.19: Imagem da interface do Mailbox.

3.3.7 Boomerang for Gmail

O Boomerang for Gmail¹⁷ é um serviço que adiciona diversas funcionalidades ao serviço de e-mail por defeito do Gmail. Disponível em versão web e em versão para Android¹⁸, pelo menos atualmente, o serviço oferece funcionalidades como a de escrever um e-mail e agendar a sua entrega para mais tarde, lembretes, seguir se um dado e-mail obteve ou não resposta e ainda a de saber se um e-mail enviado foi ou não visualizado. A grande desvantagem deste serviço prende-se com o facto de o mesmo apenas funcionar com o provedor de e-mail Gmail. O Boomerang for Gmail é um serviço com modalidade gratuita (bastante limitada) e paga, e a sua interface pode ser consultada na [Figura 3.20](#).

¹⁷<http://www.boomeranggmail.com>

¹⁸<https://www.android.com/>

E-mail

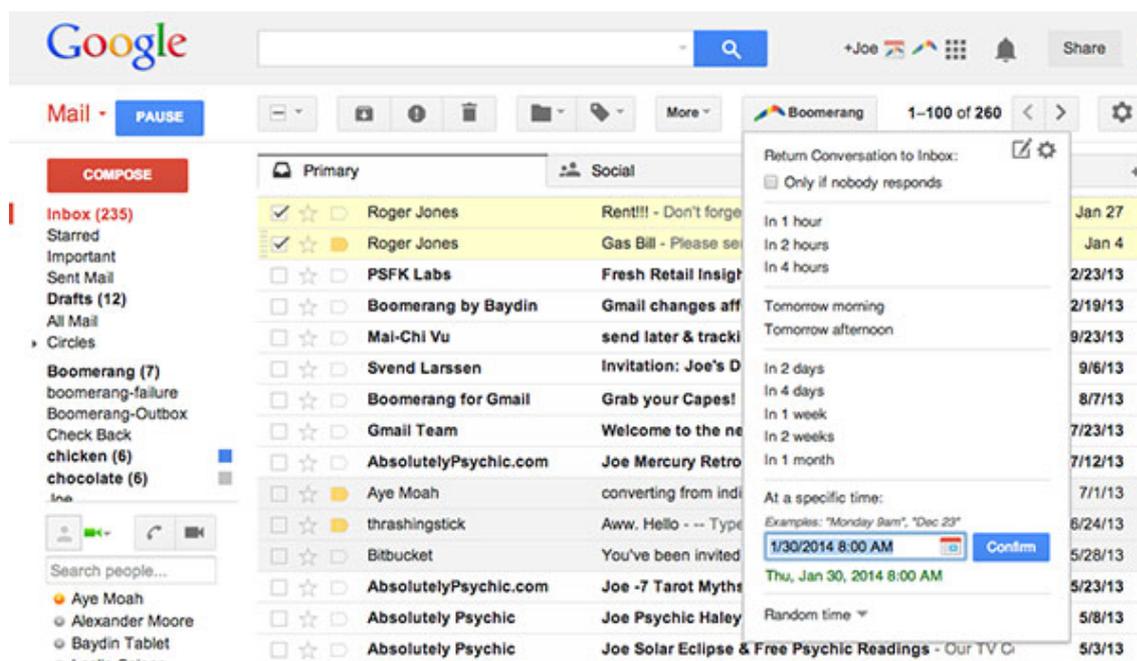


Figura 3.20: Imagem da interface do Boomerang for Gmail.

3.3.8 Análise Comparativa

Como foi possível observar anteriormente, os trabalhos industriais são vários, e são compostos por uma grande diversidade de funcionalidades e características. Para melhor se poder comparar as soluções analisadas, apresenta-se de seguida a [Tabela 3.4](#) que apresenta um comparativo entre elas, e que permitirá uma análise mais rápida e facilitada às mesmas.

Nome	Autor	Plataforma	Custo	Tipo	Gestão de Tarefas	Diferente Visualização para o E-mail	Aprendizagem Centrada no Utilizador
SaneBox	SaneBox, Inc.	Web para IMAP	Pago	Serviço	Não	Não	Sim
ActiveInbox	ActiveInbox	Web para Gmail	Pago	Plugin	Sim	Não	Não
Inboxcube	Inboxcube Inc.	iOS	Grátis	Cliente	Não	Sim	Não
Inbox by Gmail	Google	Web, Android e iOS	Grátis	Cliente	Sim	Sim	Sim
Alto by AOL	AOL	Web	Grátis	Cliente	Não	Sim	Não
Mailbox	Dropbox, Inc.	Android, iOS e OSX	Grátis	Cliente	Sim	Sim	Não
Boomerang for Gmail	Baydin, Inc.	Web para Gmail	Grátis/Pago	Plugin	Não	Não	Não

Tabela 3.4: Análise comparativa dos trabalhos industriais.

Capítulo 4

Inquérito sobre o e-mail aos utilizadores

O e-mail é uma ferramenta bastante complexa, utilizada por variadas pessoas de diferentes formas, o que faz com que seja bastante difícil perceber as alterações ideais a introduzir para o melhorar. Desta forma, para esta dissertação poder representar uma solução efetivamente capaz de aumentar a eficiência dos utilizadores no e-mail, totalmente de encontro ao que estes procuram, foi produzido um inquérito ([Apêndice A](#)) com o objetivo de perceber as necessidades, dificuldades, preferências e hábitos dos futuros participantes.

Este capítulo apresenta portanto, a metodologia utilizada para criar o questionário ([Secção 4.1](#)), os resultados obtidos do inquérito ([Secção 4.2](#)), uma discussão cuidada desses resultados ([Secção 4.3](#)) e por fim um pequeno resumo daquilo que é mais importante reter deste estudo ([Secção 4.4](#)).

4.1 Metodologia

Para ser possível levar a cabo este estudo com sucesso foi necessário ter em conta duas partes distintas: a elaboração e a disseminação. Na elaboração do questionário o mais importante foi garantir que seria possível produzir questões capazes de avaliar o necessário. Por sua vez na disseminação, o mais importante foi fazer chegar o questionário ao maior número de pessoas possível.

A dissertação apresenta como objetivo principal aumentar a eficiência dos utilizadores no e-mail, utilizando para esse fim novos métodos de visualização e algoritmos de aprendizagem. Portanto, a elaboração do questionário centrou-se em 5 categorias de perguntas:

Geral Serviu para sintetizar dados acerca dos participantes tais como género e ocupação profissional.

E-mail Serviu para reunir informação acerca dos hábitos dos participantes tais como: o local de consulta do e-mail e a frequência de consulta diária do mesmo.

Visualização Serviu essencialmente para apurar o nível de satisfação dos utilizadores com os padrões de visualização dos clientes de e-mail atuais.

Inteligência Serviu essencialmente para verificar o nível de satisfação dos utilizadores com as funcionalidades de aprendizagem dos clientes de e-mail comuns.

Considerações Finais Serviu para, através de 2 questões de resposta aberta, perceber os problemas que os utilizadores apontam às soluções de e-mail atuais, assim como para perceber o que gostariam de ver implementado numa nova solução.

Com a estrutura do inquérito bem definida, de acordo com o que era necessário avaliar e com todas as perguntas delineadas, foi necessário encontrar a estratégia de disseminação ideal. A estratégia adotada consistiu num questionário web realizado no Google Forms ¹ que foi depois enviado para toda a Universidade do Porto (alunos, docentes e não docentes) através das funcionalidades de e-mail dinâmico presentes no Sigarra ² de cada faculdade.

É ainda de referir que, no teste piloto efetuado ao questionário, o tempo médio de resposta ao questionário foi de aproximadamente três minutos.

4.2 Resultados

O inquérito foi enviado para as catorze faculdades da Universidade do Porto e chegou a cerca de 35000 pessoas (estudantes, docentes e não docentes). No final obtiveram-se 1750 participantes.

Este questionário teve como principal objetivo perceber como os utilizadores utilizavam e organizavam o seu e-mail, assim como perceber quais as funcionalidades que estes mais usavam, aquelas que estavam em excesso, as que faltavam, e qual a sua opinião à introdução das novas funcionalidades que compõe a solução proposta por esta dissertação (Capítulo 5).

Os resultados deste inquérito serviram por isso para validar e refinar as ideias delineadas para a solução proposta, e garantir que a solução final ia de encontro às expectativas dos utilizadores, que nos foram possíveis apurar. Segue-se agora a apresentação desses resultados.

4.2.1 Geral

Esta primeira categoria de perguntas do questionário fornece alguns dados demográficos sobre os participantes. Este tipo de perguntas tem bastante utilidade para garantir que inquirimos um conjunto de participantes com interesse para o estudo em questão.

A primeira observação que se pode retirar deste conjunto de dados é a da distribuição de participantes por género (Figura 4.1). Segundo os dados recolhidos 59.2% dos inquiridos eram do sexo feminino, e 40.8% do sexo masculino. Estes valores são importantes para garantir que o estudo teve participações bastante similares entre homens e mulheres.

¹<https://www.google.com/forms/about/>

²<https://sigarra.up.pt/>

Inquérito sobre o e-mail aos utilizadores

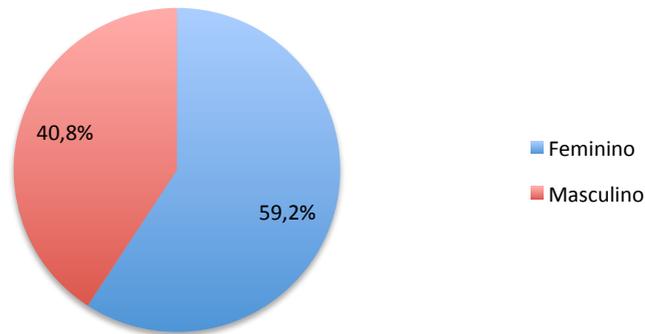


Figura 4.1: Género.

É ainda possível constatar que pelo menos 39% dos inquiridos desempenhava alguma atividade profissional (Figura 4.2), fosse ela a tempo inteiro ou a tempo parcial. Esta observação é também muito importante para assegurar que o estudo não refletia apenas resultados provenientes de um meio único.

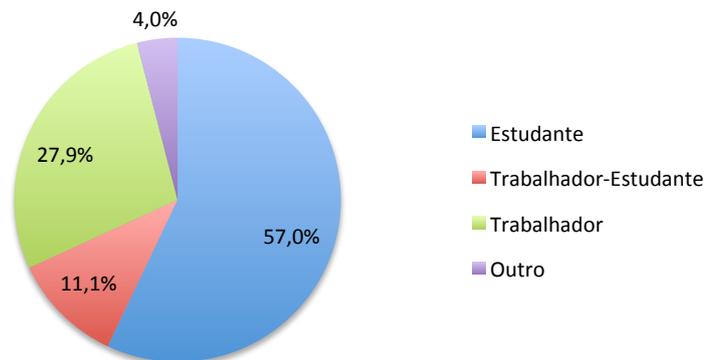


Figura 4.2: Ocupação profissional.

4.2.2 E-mail

Esta categoria do questionário, com perguntas especificamente dedicadas a aspetos gerais do e-mail, permitiu adquirir dados referentes ao perfil dos participantes enquanto utilizadores de e-mail. O tipo de cliente de e-mail utilizado, a frequência de consulta do e-mail ou o tempo médio diário despendido no e-mail são apenas algumas dessas informações.

A primeira observação a concluir é que o tipo de cliente de e-mail utilizado de forma predominante pelos utilizadores é o implementado em versões web com 62% (Figura 4.3), logo seguido pelos clientes de e-mail em versão *mobile* com 25%, menos de metade do anterior. Posto isto, como seria de esperar, o cliente de e-mail mais utilizado é o Gmail Web (um cliente de e-mail em versão web) com 64.5% dos participantes a referir que usa este cliente para ler o seu correio

Inquérito sobre o e-mail aos utilizadores

eletrónico (Figura 4.4). Seguem-se logo depois o Gmail Mobile e o Outlook Web ambos com 37.2%.

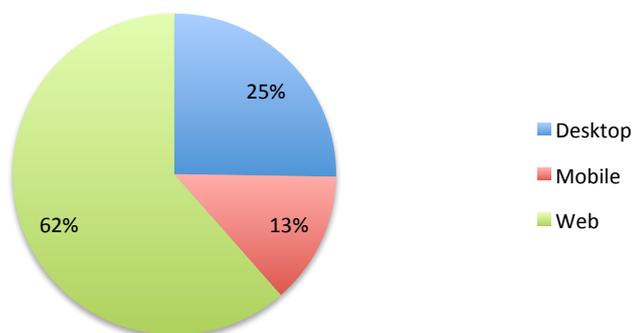


Figura 4.3: Tipo de cliente de e-mail utilizado.

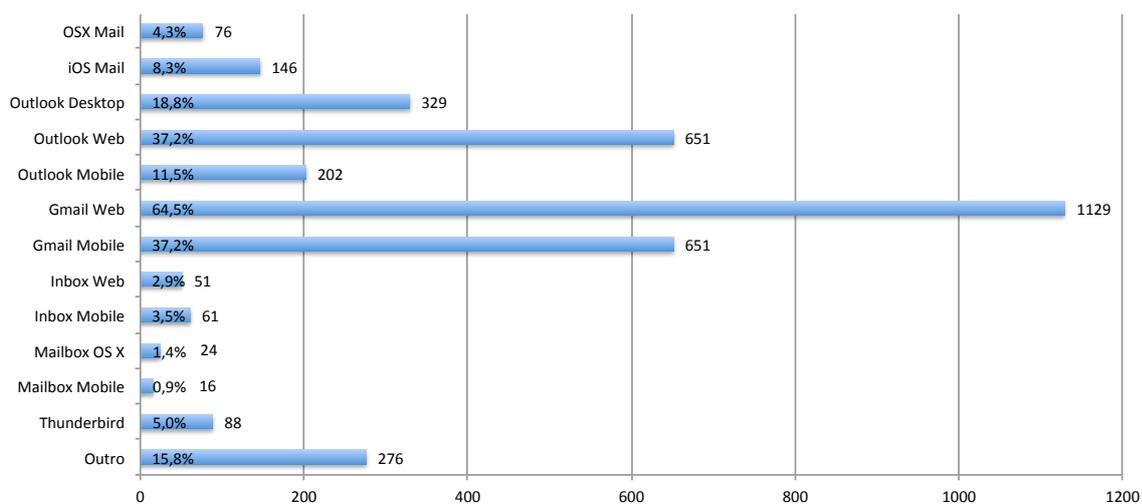


Figura 4.4: Clientes de e-mail utilizados.

A seguinte observação é também ela bastante interessante. A grande maioria dos utilizadores (78.3%) consulta o correio eletrónico em diferentes locais de forma simultânea, nomeadamente em casa ou na faculdade/escritório (Figura 4.5). Este facto denota claramente a necessidade de uma funcionalidade que permita ao utilizador separar contextualmente o tipo de e-mails que recebe. A acrescentar a este facto, é ainda de notar que uma parte muito considerável dos utilizadores (24.1%) consulta o correio eletrónico uma ou mais vezes por hora (Figura 4.6) e que 18,1% dos inquiridos depende mais de uma hora no e-mail (Figura 4.7).

Inquérito sobre o e-mail aos utilizadores

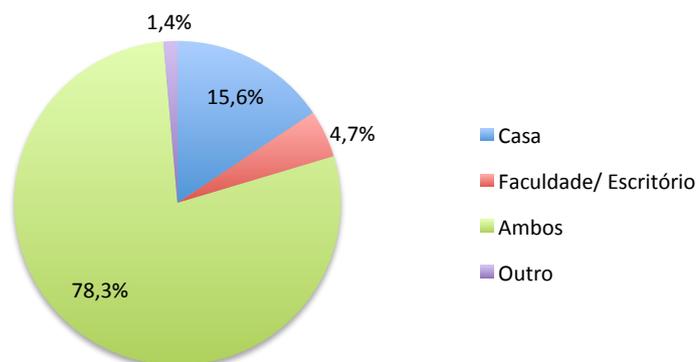


Figura 4.5: Local de consulta do e-mail.

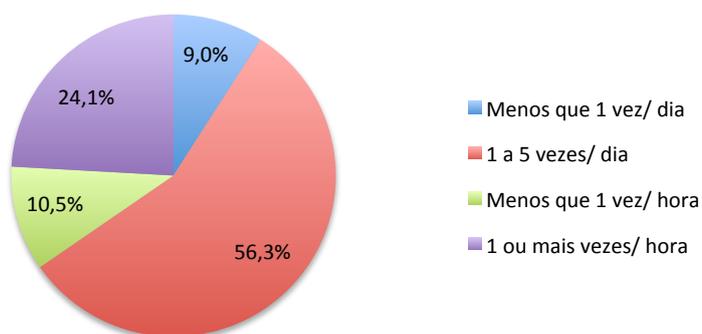


Figura 4.6: Frequência de consulta do e-mail.

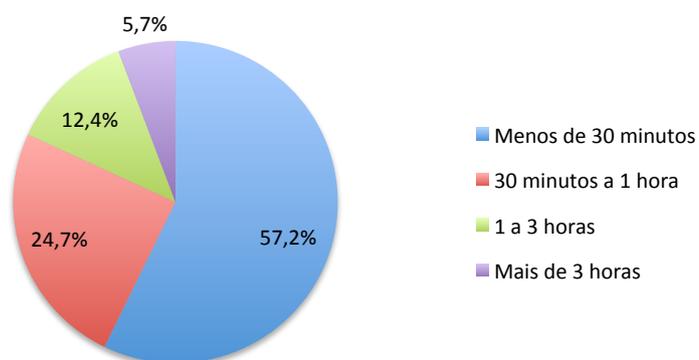


Figura 4.7: Tempo médio diário despendido no e-mail.

Por fim foi ainda possível constatar que 21% dos inquiridos recebe mais de vinte e uma mensagens de e-mail diariamente (Figura 4.8) e que 20.1% envia mais de dez mensagens de e-mail por

Inquérito sobre o e-mail aos utilizadores

dia (Figura 4.9). Se tivermos ainda em conta que 6.6% dos participantes recebe mais de cinquenta mensagens por dia, a necessidade de uma interface de e-mail eficiente torna-se evidente.

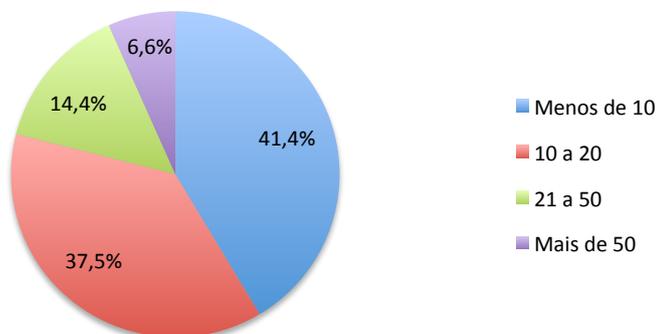


Figura 4.8: Mensagens de e-mail recebidas por dia.

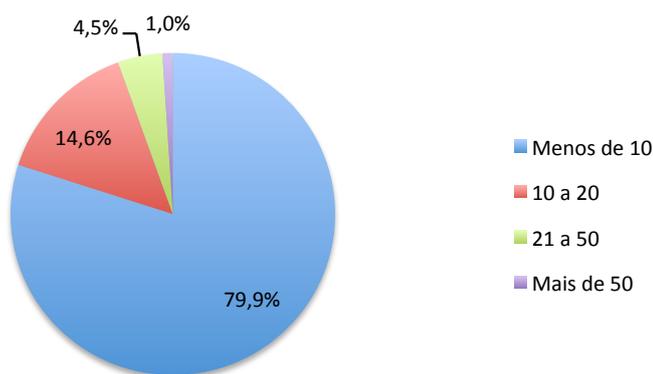


Figura 4.9: Mensagens de e-mail enviadas por dia.

4.2.3 Visualização

A próxima categoria tratada no questionário foi a da visualização. Nesta categoria foi possível adquirir dados bastante importantes sobre o modo como os utilizadores interagem com a interface do seu cliente de e-mail habitual. Tais informações permitiram mitigar dúvidas ainda existentes na solução proposta e confirmar algumas ideias já bem delineadas.

Os primeiros dois gráficos a serem analisados traduzem os resultados para o número médio de mensagens não lidas na caixa de entrada dos utilizadores (Figura 4.10) e o tipo de mensagens que os utilizadores normalmente deixam armazenadas na caixa de entrada (Figura 4.11). No primeiro deles é possível observar que 58.9% dos utilizadores tem em média menos de 10 mensagens não lidas e 25.1% tem em média de 10 a 50 mensagens não lidas. Já no segundo gráfico é possível observar que os utilizadores deixam armazenadas, na caixa de entrada, mensagens com conversas

Inquérito sobre o e-mail aos utilizadores

em curso, mensagens não arquivadas e tarefas incompletas, sendo que as percentagens são 69%, 43.2% e 53.7% respetivamente.

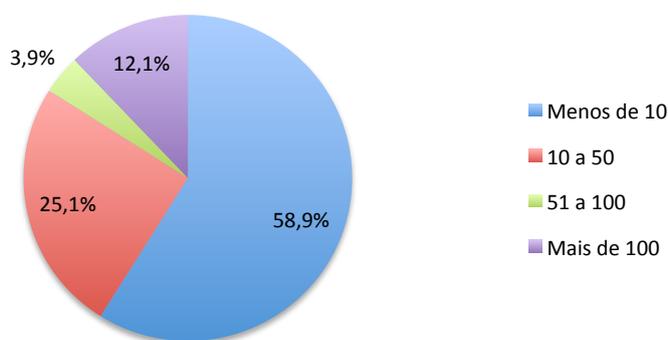


Figura 4.10: Número médio de mensagens não lidas na caixa de entrada.

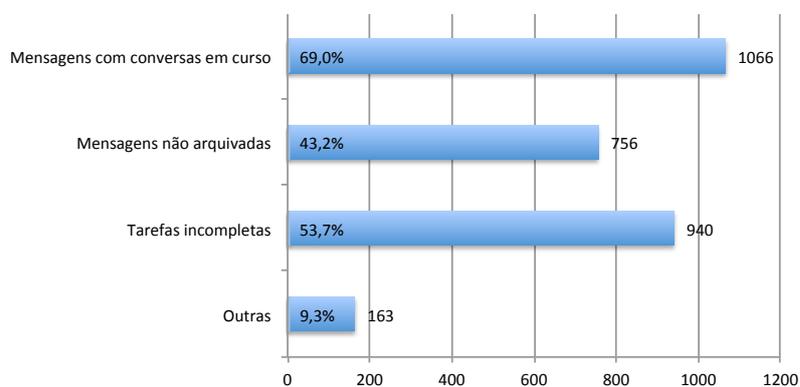


Figura 4.11: Tipo de mensagens de e-mail deixadas na caixa de entrada.

Continuando com a nossa análise é possível observar mais três gráficos importantes. O primeiro resume as estatísticas recolhidas acerca do número de pastas presentes no correio eletrónico dos utilizadores (Figura 4.12), excluindo claro está as já criadas por defeito. Consta-se então que 38.2% dos utilizadores usa menos de 2 pastas extra e que 25.3% utiliza de 2 a 5, ou seja, 63.5% dos utilizadores utiliza menos que 5 pastas no seu e-mail. O segundo gráfico trata de analisar o método de pesquisa de mensagens mais utilizado (Figura 4.13), sendo que, a esmagadora maioria (73.5%) utiliza a barra de pesquisa. Por último, o terceiro gráfico apresenta informação sobre o critério de pesquisa mais utilizado (Figura 4.14), e como é possível observar, 57.9% dos inquiridos utiliza predominantemente o remetente.

Inquérito sobre o e-mail aos utilizadores

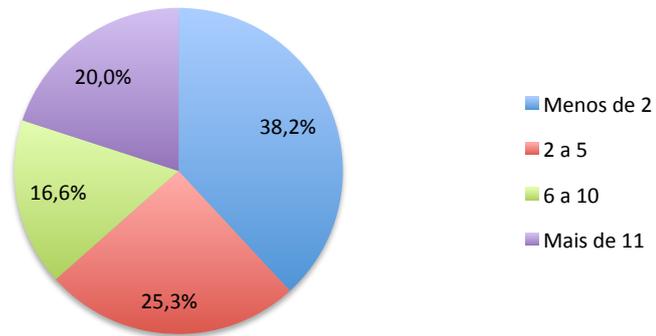


Figura 4.12: Número de pastas excluindo as existentes por defeito.

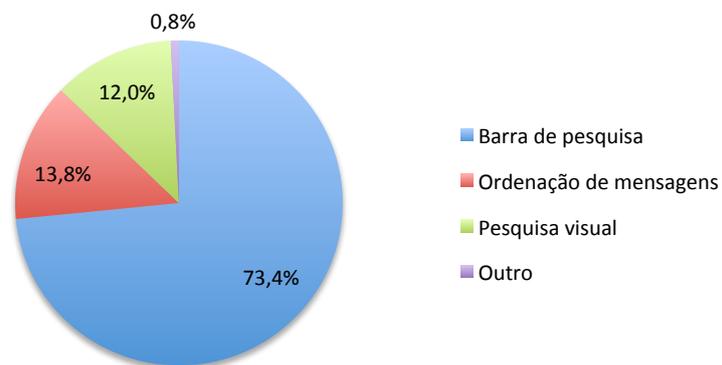


Figura 4.13: Método de pesquisa de mensagens mais utilizado.

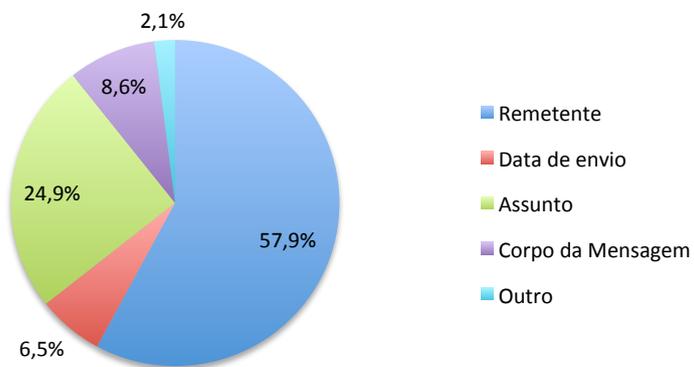


Figura 4.14: Critério de pesquisa mais utilizado.

Inquérito sobre o e-mail aos utilizadores

De seguida partiremos para a análise de dois gráficos que resumem a opinião dos participantes em relação às notificações automáticas. O primeiro regista quais os tipos de notificações automáticas utilizadas (Figura 4.15), sendo que 56.6% utiliza notificações automáticas de receção em pelo menos numa das suas vertentes (auditiva ou visual). O segundo regista se os inquiridos consideram, ou não, as notificações automáticas intrusivas (Figura 4.16) ao que 70.7% responde negativamente.

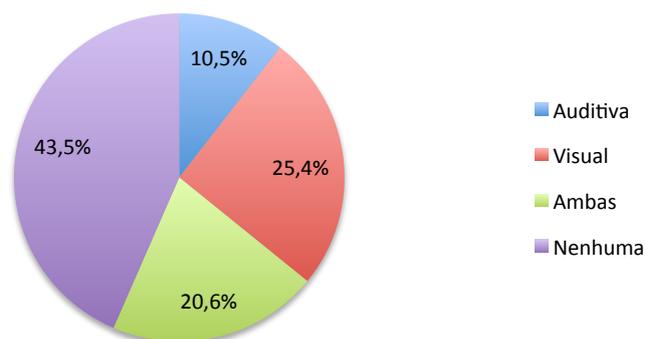


Figura 4.15: Tipo de notificações automáticas utilizadas.

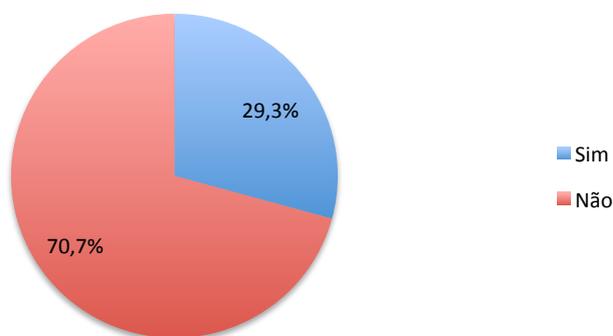


Figura 4.16: Classificação das notificações automáticas quanto à sua intrusividade.

Por fim, nesta categoria do questionário, apresentar-se-ão mais três gráficos. O primeiro deles retrata a frequência com que os participantes, no processo de criação de uma nova mensagem, utilizam uma já recebida anteriormente para obter o e-mail do destinatário que pretendem (Figura 4.17). Os resultados apontam que 21.7% considera esta prática muito frequente e que 29.3% a considera frequente. Os segundo e terceiro gráficos tentam avaliar, respetivamente, o nível de satisfação dos utilizadores face à atual interface dos clientes de e-mail (Figura 4.18) e a opinião dos utilizadores quanto à prejudicialidade da atual visualização do e-mail face à sua produtividade pessoal (Figura 4.19). Face aos resultados constata-se que 14.3% dos inquiridos se encontra muito

Inquérito sobre o e-mail aos utilizadores

satisfeito e 59.1% satisfeito face à atual interface dos clientes de e-mail, e ainda que 70.7% não considera a atual visualização prejudicial à sua produtividade pessoal.

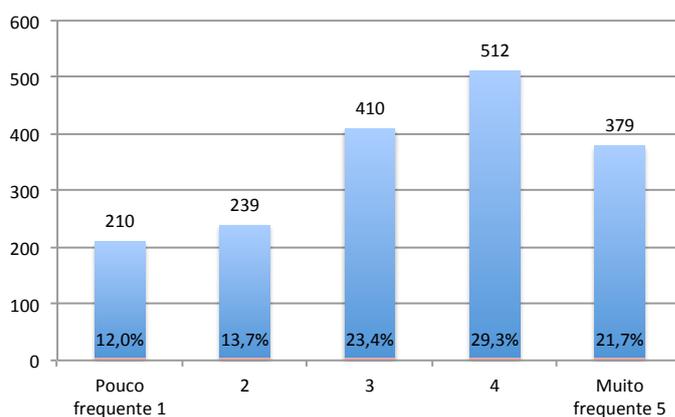


Figura 4.17: Frequência de utilização de uma mensagem já recebida para obter o e-mail do destinatário para uma nova mensagem. $\bar{x} \approx 3.35$; $\sigma \approx 1.29$

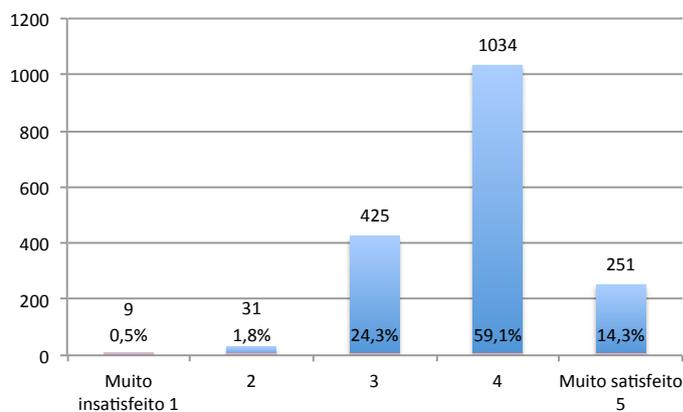


Figura 4.18: Satisfação dos utilizadores em relação à atual interface dos clientes de e-mail. $\bar{x} \approx 3.85$; $\sigma \approx 0.69$

Inquérito sobre o e-mail aos utilizadores

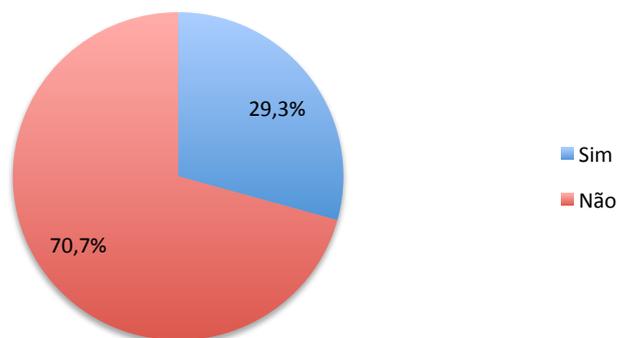


Figura 4.19: Opinião dos utilizadores quanto à prejudicialidade da atual visualização do e-mail face a sua produtividade pessoal.

4.2.4 Inteligência

A categoria tratada de seguida no questionário foi a da inteligência. Nesta categoria foi possível adquirir dados sobre a opinião que os utilizadores têm sobre as funcionalidades de aprendizagem e inteligência artificial já presentes nos clientes de e-mail atuais. Tais informações permitiram mais uma vez eliminar algumas dúvidas, ainda existentes, referentes a esta temática na solução proposta, e confirmar ainda algumas ideias já bem delineadas.

Os primeiros três gráficos a serem analisados nesta categoria tratam o nível de satisfação com as funcionalidades de inteligência e aprendizagem já desenvolvidas para o e-mail (Figura 4.20), a receção de e-mails sem interesse em serem lidos aquando da sua chegada (Figura 4.21) e por fim a opinião dos utilizadores sobre a sobrecarga causada pela quantidade de mensagens de e-mail a processar (Figura 4.22). Com recurso aos dados visíveis, é possível observar que 76.1% dos inquiridos diz-se satisfeito com as funcionalidades de inteligência e aprendizagem já desenvolvidas, que 87.2% sentem que recebem e-mails que não têm interesse em serem lidos aquando da sua chegada, e por fim que 51.8% não se sente sobrecarregado pela quantidade de mensagens de e-mail que tem para processar.

Inquérito sobre o e-mail aos utilizadores

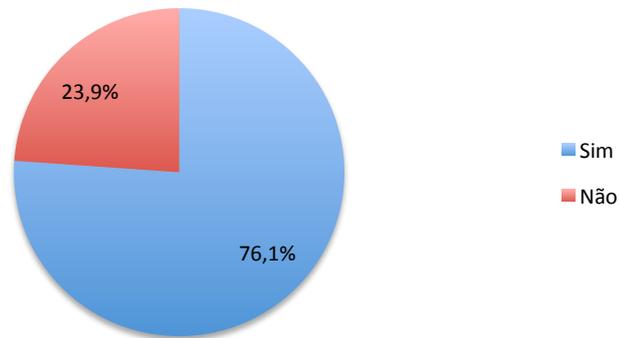


Figura 4.20: Nível de satisfação com as funcionalidades de inteligência e aprendizagem já desenvolvidas para o e-mail.

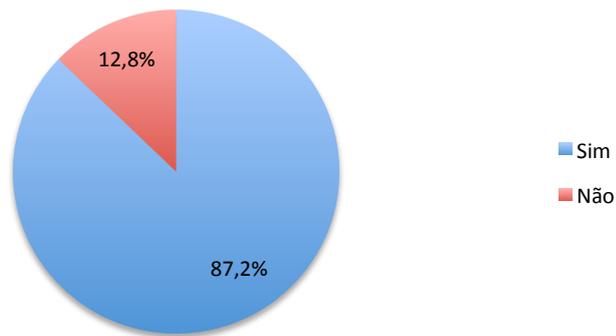


Figura 4.21: Receção de e-mails sem interesse em ser lidos aquando da sua chegada.

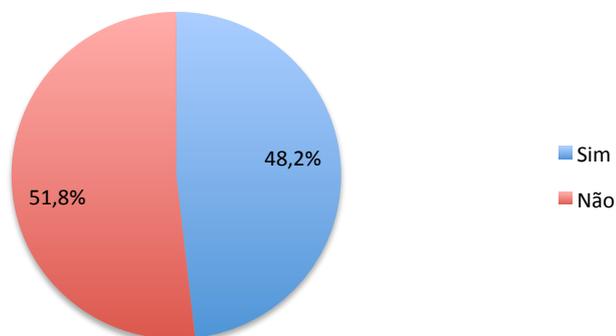


Figura 4.22: Opinião sobre sobrecarga causada pela quantidade de mensagens de e-mail para processar.

Inquérito sobre o e-mail aos utilizadores

Por fim os últimos três gráficos existentes, para analisar esta categoria, resumem informação sobre a pretensão por um cliente de e-mail capaz de se adaptar ao perfil de cada utilizador (Figura 4.23), sobre a pretensão por perceber o conteúdo de um conjunto de mensagens através da visualização de meta-informação (Figura 4.24) e resumem informação sobre a opinião dos participantes quanto ao aumento de produtividade e eficiência causado por uma classificação das *threads* de e-mail em trabalho e lazer (Figura 4.25). É de constatar então, depois de visualizados os gráficos, que 74.9% dos participantes pretendem um cliente de e-mail que se adapte ao seu perfil pessoal, 87.9% pretendem perceber o conteúdo das mensagens antes mesmo de as abrir e que 89.9% considera que se verificaria um aumento na sua produtividade e eficiência pessoal se usassem um cliente de e-mail capaz de classificar as *threads* de e-mail em trabalho ou lazer.

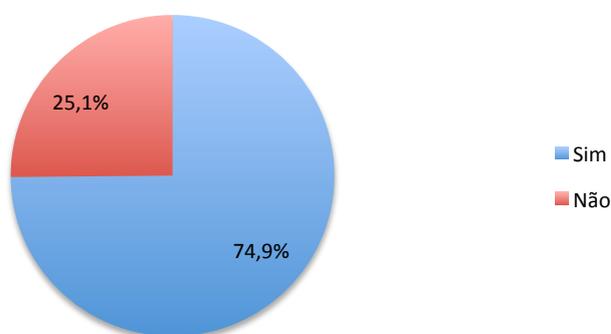


Figura 4.23: Pretensão por um cliente de e-mail capaz de se adaptar ao perfil do utilizador.

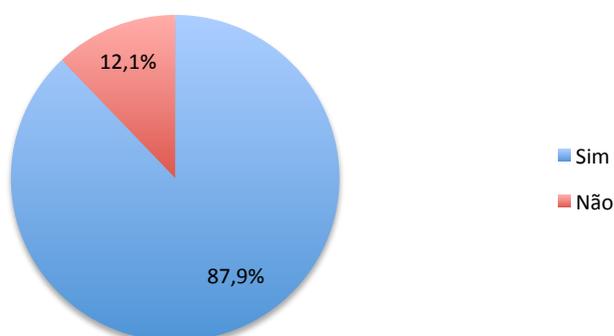


Figura 4.24: Pretensão por perceber o conteúdo de um conjunto de mensagens antes mesmo de as ler, através da visualização de meta-informação.

Inquérito sobre o e-mail aos utilizadores

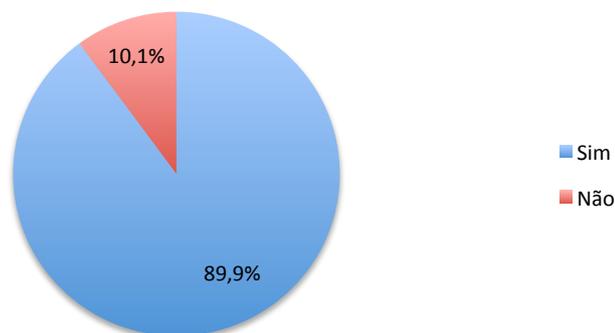


Figura 4.25: Opinião sobre aumento de produtividade e eficiência no uso do e-mail, causado por uma classificação clara das *threads* em trabalho ou lazer.

4.2.5 Considerações Finais

O questionário efetuado terminava com duas questões de resposta aberta, onde os utilizadores tinham a liberdade de apontar quais os problemas por eles encontrados nos clientes de e-mail atuais, e quais as sugestões de novas funcionalidades que tinham para um novo cliente de e-mail. Os resultados obtidos foram de todo muito interessantes e encontram-se resumidos, e agrupados, na [Tabela 4.1](#) e na [Tabela 4.2](#) apresentadas de seguida.

Problemas encontrados nos clientes de e-mail atuais	
Filtros pouco versáteis e eficazes.	Impossibilidade de envio programado de mensagens.
Notificações pouco personalizáveis.	Deficiente gestão de contactos.
Interface de resposta e leitura de mensagem dependente.	Processo para arquivar e-mails demorado.
Confirmação de receção.	Visualização de conversas dentro de <i>threads</i> pouco intuitiva.
Filtros de <i>spam</i> pouco eficazes.	Pouca flexibilidade na organização da interface.
Difícil organização de conversas.	Funcionalidades para gestão de tarefas pouco desenvolvidas.
Falta de divisão entre mensagens de trabalho e mensagens de lazer.	Pesquisa pouco eficaz.
Tamanho máximo para anexos demasiado pequeno.	Falta de privacidade.
Deficiente classificação das mensagens de e-mail.	Difícil gestão de pastas.
Interface do e-mail não orientada ao remetente.	Falta da funcionalidade de bloquear remetente.
Pesquisa por remetente é efetuada pela barra de pesquisa sendo necessário uma melhor maneira de o fazer.	Processo de tratamento de mensagens demasiado dispendioso quando a afluência de mensagens é elevada.
Falta de inteligência artificial no e-mail (organização automática, hierarquia automática e lembretes para resposta a mensagens importantes, já lidas mas sem resposta).	

Tabela 4.1: Lista dos problemas encontrados pelos inquiridos nos clientes de e-mail atuais.

Inquérito sobre o e-mail aos utilizadores

Funcionalidades pretendidas num novo cliente de e-mail	
Agrupamento por temas ou grupos de discussão.	Funcionalidade de bloquear remetente.
Priorização da caixa de entrada pela urgência do assunto.	Envio programado de mensagens de e-mail.
Filtragem eficaz.	Gestão de tarefas.
Classificação de mensagens em trabalho ou lazer.	Gestão de contactos.
Calendário semanal visível na caixa de entrada.	Vista centralizada para os documentos e anexos.
Resumo de mensagens de e-mail.	Aviso de entrega / Confirmação de receção.
Leitura, com recurso a voz, dos e-mails.	Filtro de <i>spam</i> eficaz.
Funcionalidade de “Ler mais tarde”.	Cancelamento do envio de mensagens acidentais.
Interface de resposta a uma mensagem semelhante à de um cliente de mensagens instantâneas.	Filtros automáticos mediante o assunto ou palavras-chave.
Estatísticas acerca das mensagens enviadas (temas e destinatários de e-mails aos quais o utilizador responde).	Funcionalidade de sugestão de assunto aquando da criação de uma nova mensagem.
Agrupamento das <i>newsletters</i> ou outras mensagens não importantes.	Agrupamento das mensagens de e-mail por remetente como numa aplicação de conversação.
Emissão de notificações apenas para mensagens importantes.	

Tabela 4.2: Lista das funcionalidades pretendidas pelos inquiridos num novo cliente de e-mail.

4.3 Discussão dos Resultados

A presente discussão de resultados servirá essencialmente para retirar as devidas conclusões dos resultados apresentados anteriormente. Os resultados já apresentados influenciaram, de forma significativa, a solução apresentada no próximo capítulo.

Como já observado, a participação no questionário de homens e mulheres foi bastante similar. Tal facto permite, de certa forma, afirmar que esta variável não afetou significativamente os resultados apresentados, o que se torna bastante importante. O mesmo se passa para a ocupação profissional dos inquiridos: registou-se que pelo menos 39% desempenhava alguma atividade profissional, o que também permite considerar que este estudo não apresenta apenas as necessidades de um determinado meio.

Continuando a análise é possível concluir também que os participantes utilizam o e-mail maioritariamente na sua componente web (62%), e que o cliente de e-mail que regista maior utilização é o Gmail Web (64.5%). Estes resultados traduzem a ideia de que, a desenvolver uma versão de um novo cliente de e-mail, o mesmo deve conter obrigatoriamente uma versão web. A adicionar ainda ao concluído, observa-se que 78.3% dos inquiridos consulta o e-mail simultaneamente em casa e na faculdade/escritório, e tais resultados, sugerem claramente que poderá haver a necessidade de os utilizadores conseguirem distinguir, segundo o contexto, quais as mensagens de e-mail importantes para si em cada um desses diferentes momentos. Os valores registados para a frequência de consulta do e-mail, para o tempo médio diário despendido, para as mensagens recebidas e enviadas diariamente constituem também importantes elementos para demonstrar que os participantes

Inquérito sobre o e-mail aos utilizadores

despendem bastante tempo do seu dia a utilizar o e-mail. Cerca de 37% dos utilizadores consulta o e-mail pelo menos 1 vez por hora, aproximadamente 18% afirma despende pelo menos 1 hora diária a utilizá-lo, 21% recebem pelo menos 21 mensagens por dia e 19% envia pelo menos mais de 10 mensagens. Estes últimos dados, para além de refletirem o elevado uso do e-mail, refletem também que melhorar a interface envolvida na receção e apresentação das mensagens de e-mail ao utilizador poderá causar maior impacto na qualidade final da solução.

Prosseguindo a análise, para a categoria do questionário da visualização, é possível concluir que os utilizadores poderão estar com alguma dificuldade em gerir a sua caixa de entrada, isto porque em média cerca de 41% dos utilizadores tem pelo menos 10 mensagens de e-mail não lidas na sua caixa de entrada. Foi possível também concluir que para os utilizadores a funcionalidade de pastas não é algo verdadeiramente fulcral, e pode ser aproximada com recurso a outras funcionalidades que diminuem a complexidade da interface, por exemplo etiquetas, uma vez que cerca de 64% dos inquiridos utiliza 5 ou menos pastas no seu correio eletrónico. A barra de pesquisa assumiu-se como um padrão, no caso de os utilizadores precisarem de efetuar a pesquisa por uma mensagem, uma vez que 73% afirmou ser o método de pesquisa que utiliza mais vezes. Já no que toca ao critério de pesquisa concluiu-se que o mais utilizado é o remetente, com aproximadamente 58% dos inquiridos a afirmá-lo. As notificações automáticas, por sua vez, são uma funcionalidade bastante utilizada. Cerca de 69% dos participantes utiliza pelo menos uma das suas vertentes (visual e auditiva) e aproximadamente 70% afirma não as considerar intrusivas. Uma outra informação muito relevante, dado já ter sido visto anteriormente que o remetente é o critério de pesquisa mais utilizado, é a de 51% dos inquiridos, quando pretende enviar uma nova mensagem, considera pelo menos frequente o facto de procurar uma mensagem já recebida da pessoa a quem pretende enviar a nova mensagem, para assim proceder com o envio. É de concluir portanto que orientar a visualização do cliente de e-mail ao remetente, ao invés de orientar ao assunto, poderá ser vantajoso. Por fim conclui-se ainda que, pese embora o facto de cerca de 59% dos inquiridos se dizer satisfeito com a atual interface dos clientes de e-mail, existem ainda cerca de 29% dos inquiridos que consideram que a atual visualização do e-mail está a prejudicar a sua produtividade pessoal, o que denota um problema que é necessário resolver.

A categoria de inteligência, presente no questionário, foi também ela bastante rica em conclusões. Uma percentagem considerável dos participantes (cerca de 24%) considera que não se encontra satisfeito com as funcionalidades de inteligência e aprendizagem já desenvolvidas para o e-mail, e uma percentagem ainda maior (87%) afirma que recebe e-mails sem importância em serem lidos aquando da sua chegada. Outros 48% dizem ainda se sentirem sobrecarregados pela quantidade de e-mails que têm para processar. É ainda de concluir que os utilizadores pretendem uma solução integrada de e-mail que faça com que este seja capaz de se adaptar ao perfil de cada utilizador (75%), que permita perceber o conteúdo de um dado conjunto de mensagens antes mesmo de as abrir (88%) e que classifique ainda, de forma automática, as *threads* de e-mail em trabalho ou lazer (90%).

4.4 Resumo

Por fim, para terminar este capítulo, apresentar-se-á um resumo, sob a forma de listagem, de toda a informação importante referida anteriormente ao longo deste capítulo. Com o referido resumo pretende-se fornecer ao leitor, uma forma rápida de entender os pontos fulcrais deste estudo realizado à Universidade do Porto.

- Participação aproximada de homens e mulheres.
- Participação aproximada entre trabalhadores e estudantes.
- A grande maioria dos utilizadores (62%) utiliza o cliente de e-mail no *browser*, o que indica a necessidade de um novo cliente de e-mail ter de contar com uma versão web.
- A grande percentagem de utilizadores que consulta o e-mail, de forma simultânea, em casa e no seu local de trabalho, sugere que poderá existir uma necessidade em o e-mail perceber a importância de uma dada mensagem de e-mail em cada um desses diferentes contextos.
- Cerca de 37% dos utilizadores consulta o e-mail pelo menos 1 vez por hora e aproximadamente 18% afirma despende pelo menos 1 hora diária a utilizá-lo, 21% recebem pelo menos 21 mensagens por dia e 19% envia pelo menos mais de 10 mensagens. Tais dados refletem que o e-mail é bastante utilizado, e que as alterações na interface de receção e apresentação de mensagens dos clientes de e-mail poderão causar um maior impacto na qualidade da solução final.
- 41% dos utilizadores tem pelo menos 10 mensagens de e-mail não lidas na sua caixa de entrada, o que denota uma dificuldade dos mesmos em gerir a sua caixa de correio.
- 64% dos inquiridos utiliza 5 ou menos pastas no seu correio eletrónico, o que poderá traduzir que esta funcionalidade não é verdadeiramente fulcral para os utilizadores, e pode ser aproximada com recurso a outras funcionalidades que diminuem a complexidade da interface, como por exemplo etiquetas.
- A barra de pesquisa é o método de pesquisa mais utilizado (73%) e o remetente o critério de pesquisa preferido dos utilizadores (58%).
- As notificações automáticas de receção afirmam-se como bastante importantes pois 69% utiliza pelo menos uma das suas vertentes (visual e auditiva) e 70% afirma não as considerar intrusivas.
- Dado que o remetente é o critério de pesquisa mais popular e que 51% dos participantes considera pelo menos frequente utilizar uma mensagem já recebida para obter o e-mail do destinatário para uma nova mensagem, é concluído que poderá ser vantajoso orientar a visualização do e-mail ao remetente, ao invés de a ter orientada ao assunto.

Inquérito sobre o e-mail aos utilizadores

- Existe um problema com a interface de e-mail que é necessário resolver, uma vez que cerca de 29% dos participantes afirma que a atual visualização está a afetar a sua produtividade pessoal.
- Existe um problema com a inteligência incluída no e-mail, uma vez que cerca de 24% considera que não se encontra satisfeito com as funcionalidades de inteligência e aprendizagem já desenvolvidas para o e-mail, 87% afirma que recebe e-mails sem importância em ser lidos aquando da sua chegada e 48% dizem ainda se sentirem sobrecarregados pela quantidade de e-mails que têm para processar.
- Os utilizadores pretendem uma solução integrada de e-mail que faça com que este seja capaz de se adaptar ao perfil de cada utilizador (75%), que permita perceber o conteúdo de um dado conjunto de mensagens antes mesmo de as abrir (88%) e que classifique ainda, de forma automática, as *threads* de e-mail em trabalho ou lazer (90%).

Capítulo 5

Spidious: Solução

O presente capítulo pretende, em primeiro lugar, oferecer uma visão mais clara do problema que esta dissertação aborda. Em segundo lugar pretende abordar também a solução que se adotou para o mesmo. Serão analisadas a arquitetura e as tecnologias utilizadas na construção da solução, a visualização e os algoritmos de inteligência e aprendizagem presentes, e por fim será analisada toda a solução final integrada.

À solução agora proposta foi dado o nome de Spidious. A origem do nome prende-se com a junção da palavra *speedy* com o sufixo *-ous* que tem o significado de “cheio de”. Dessa junção surgiu o nome Spidious, ao qual se pretende associar a noção de rapidez.

Este capítulo resume portanto todo o trabalho inovador resultante desta dissertação. O Spidious pretende assumir-se como uma solução competente na resolução do problema discutido, melhorando os índices de eficiência dos utilizadores no e-mail.

5.1 Problema

O facto de o e-mail enfrentar um problema e estar sobrecarregado é inegável. O e-mail é o meio de comunicação eletrónica preferencial para muita gente, é utilizado para fins pessoais e profissionais, mas não foi pensado para os padrões atuais. O número de mensagens recebidas num único dia pode ser muito elevado, e facilmente mensagens muito importantes misturam-se com outras menos importantes. O resultado final de tudo isto, tende a ser uma sensação de aflição tremenda nos utilizadores que deixam de conseguir lidar com todo aquele fluxo de informação. E aqueles que o fazem de uma forma eficaz, tendem a ver a sua eficiência pessoal reduzida [JDW01].

Para além disso, o correio eletrónico foi projetado para ser uma ferramenta de comunicação assíncrona, e tornou-se bem mais que isso. E esta é de facto a primeira causa para todo este problema. As pessoas tendem a usar o e-mail para arquivo pessoal, troca de documentos, gestão de contactos e ainda para gestão de tarefas e atividades [DB01, WS96].

Desta forma, o utilizador rapidamente sentiu a necessidade de um sistema melhor que tende em não aparecer. Várias soluções foram aparecendo como as pastas, as *threads*, a pesquisa, os filtros e a ordenação, mas a verdadeira ideia que se retém é que os clientes de e-mail precisam de

ser redesenhados e repensados para suportar a grande diversidade de utilizadores e funções para o qual é usado.

A abordagem ao problema por parte desta dissertação pretende oferecer ao utilizador um cliente de e-mail web no qual ele possa facilmente, e de forma rápida, adquirir a informação mais importante contida nas suas mensagens de e-mail. A solução, na perspetiva desta dissertação, passa por oferecer ao utilizador uma nova visualização no contexto do e-mail, complementada por algoritmos de aprendizagem baseados em ações do utilizador.

5.2 Tecnologias

A solução proposta de seguida é em todos os aspetos bastante ambiciosa. Não só pretende tratar um problema difícil, como pretende resolvê-lo utilizando um meio que apresenta diversas dificuldades técnicas (uma solução na nuvem). Desta forma, as tecnologias foram escolhidas tendo em conta a facilidade com que escalam horizontalmente, e a rapidez de desenvolvimento que oferecem.

5.2.1 HTML5

O HTML5¹ (*HyperText Markup Language 5*) é uma linguagem para estruturação e apresentação de conteúdos web, e uma tecnologia chave da web desde do início. A sua quinta convenção apresenta importantes melhorias e novas funcionalidades que mantêm o HTML útil nos paradigmas de utilização atuais. A atualização mais notada na nova versão assenta no suporte nativo a conteúdos audiovisuais. O HTML5 tornou-se assim o novo padrão para HTML e XHTML e está em constante evolução, sendo que a maioria dos navegadores web já implementam grande parte das suas funcionalidades.

5.2.2 CSS3

O CSS3² (*Cascading Style Sheets 3*) é a mais recente atualização do popular CSS, um mecanismo simples que permite criar estilos para páginas web, ao mesmo tempo que permite ainda adicionar efeitos de transição, imagens e outros componentes integrantes, do que se veio a designar mais tarde por web 2.0.

A sua facilidade e potencial criativo colocam esta tecnologia na vanguarda do *webdesign*, tornando-a também numa das formas de apresentação web mais utilizadas da atualidade.

5.2.3 JavaScript

O JavaScript³ é uma linguagem de programação interpretada, que nasceu com o propósito de dar aos navegadores web a possibilidade de executarem scripts do lado do cliente, permitindo

¹<http://www.w3.org/TR/html5/>

²<http://www.w3.org/TR/css-2010/>

³<http://www.ecmascript.org>, <https://developer.mozilla.org/en-US/docs/Web/JavaScript>

dessa forma uma interação do utilizador com a página web, sem que o script tenha de passar pelo servidor.

Ainda de referir que foi concebida para tomar a forma de uma linguagem de script orientada a objetos, de tipagem fraca e dinâmica. Possui ainda suporte à programação funcional. Apresenta-se hoje em dia como uma linguagem ideal para implementação de aplicações web devido à rapidez de desenvolvimento que proporciona.

Por estas razões, esta foi a linguagem principal escolhida para o desenvolvimento da solução.

5.2.4 NodeJS

O NodeJS⁴ é uma plataforma construída sobre o motor de JavaScript do Chrome o V8⁵ para tornar fácil o desenvolvimento de aplicações de rede rápidas e escaláveis. O NodeJS usa um modelo de entradas e saídas não bloqueante orientando a eventos, que o torna extremamente leve e eficiente.

É a solução perfeita para aplicações em tempo real que movimentam muita quantidade de dados e que correm de forma distribuída. Foi por isso a tecnologia escolhida para ser utilizada como servidor para a solução.

5.2.5 EmberJS

O EmberJS⁶ tem como base o SproutCore⁷. É uma ferramenta em JavaScript que implementa o padrão de desenho MVC (*Model View Controller*) e que tem como objetivo o de permitir a criação de aplicações web que se aproximem as aplicações de computador.

O EmberJS permite essencialmente a criação de aplicações web de forma mais organizada, e permite que tarefas de manipulação do website complicadas se tornem verdadeiramente simples. Por estas razões foi a ferramenta escolhida para implementar a parte da solução que fará interface com o utilizador.

5.2.6 RethinkDB

A RethinkDB⁸ é uma base de dados orientada a documentos, bastante similar ao MongoDB⁹, mas que dá especial ênfase à escalabilidade. A RethinkDB foi construída para armazenar documentos JSON¹⁰ (*JavaScript Object Notation*) e escalar facilmente para vários servidores sem praticamente nenhum esforço por parte dos seus utilizadores. É ainda de assinalar a sua excelente linguagem de consulta, que permite efetuar consultas realmente importantes como junção de tabelas e agrupamentos. É ainda de referir que é fácil de aprender e de configurar.

⁴<http://nodejs.org>

⁵<https://code.google.com/p/v8/>

⁶<http://emberjs.com>

⁷<http://sproutcore.com>

⁸<http://rethinkdb.com>

⁹<http://www.mongodb.org>

¹⁰<http://json.org>

Por todas estas razões, este foi o sistema de base de dados escolhido para o desenvolvimento desta solução.

5.2.7 WebSocket

O WebSocket¹¹ é um protocolo que permite comunicações bidirecionais entre um cliente e um servidor, utilizando uma comunicação TCP¹² (*Transmission Control Protocol*). Esta tecnologia torna possível a criação de uma sessão de comunicação interativa entre o *browser* do utilizador e o servidor. Dessa forma, permite o uso de comunicações orientadas a eventos que eliminam por completo a necessidade de efetuar verificações para confirmar a existência de informação nova, quer do lado do servidor, quer do lado do cliente.

Dadas todas estas características, esta é uma tecnologia fulcral no desenvolvimento desta solução.

5.3 Arquitetura

A solução que se propõe nesta dissertação, ao ser materializada num cliente web que corre na nuvem, e dado o problema, tem que obrigatoriamente pensar em escalabilidade. A razão para todas estas preocupações é muito simples: se a solução se demonstrar comprovadamente boa tem que ser capaz de responder à natural afluência de utilizadores interessados em utilizá-la. Desse modo, a arquitetura proposta contempla um balanceador de carga, um conjunto de máquinas a servir a aplicação web (NodeJS e EmberJS) e um conjunto de servidores de base de dados. A arquitetura pode ser consultada na [Figura 5.1](#).

A chegada a esta estrutura não foi de todo simples. Devido à complexidade de toda a solução foram necessárias várias horas de pesquisa e de estudo de arquiteturas comprovadamente escaláveis e já em produção no mercado, tal como a do *Instagram*¹³, a do *Twitter*¹⁴, a do *Pinterest*¹⁵ e a do *Stack Overflow*¹⁶. Foram ainda consultadas outras publicações bem aceites pela comunidade como a seguinte do *GoldFire Studios*¹⁷.

No entanto todas as opções tomadas são de fácil entendimento. O balanceador de carga existe para, essencialmente, encaminhar todo o tráfego de entrada para uma máquina do *cluster* que está a servir a aplicação principal, uma vez que desta forma se consegue fazer uma divisão eficiente dos utilizadores pelos recursos disponíveis. Ao mesmo tempo, de forma paralela, dota-se a estrutura da capacidade de em qualquer momento aumentar o seu *cluster* de máquinas que servem a aplicação principal e fazê-la albergar mais utilizadores. A escalabilidade do próprio balanceador de carga também está assegurada, uma vez que não existe apenas um balanceador de carga mas

¹¹<https://www.websocket.org>

¹²<https://www.ietf.org/rfc/rfc793.txt>

¹³<http://highscalability.com/blog/2012/4/9/the-instagram-architecture-facebook-bought-for-a-cool-billio.html>

¹⁴<http://highscalability.com/scaling-twitter-making-twitter-10000-percent-faster>

¹⁵<http://highscalability.com/blog/2013/4/15/scaling-pinterest-from-0-to-10s-of-billions-of-page-views-a.html>

¹⁶<http://highscalability.com/blog/2014/7/21/stackoverflow-update-560m-pageviews-a-month-25-servers-and-i.html>

¹⁷<http://goldfirestudios.com/blog/136/Horizontally-Scaling-Node.js-and-WebSockets-with-Redis>

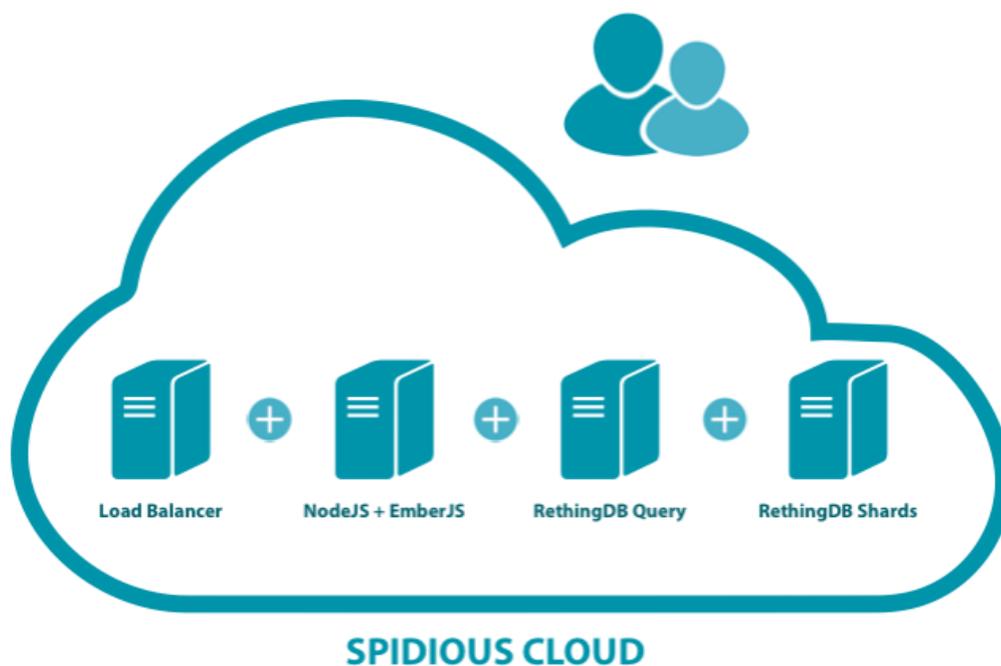


Figura 5.1: Arquitetura da solução proposta.

sim vários. O acesso aos nós de balanceamento de carga está assegurado por *DNS (Domain Name System) round robin*. Por fim, a escalabilidade da base de dados escolhida (*RethinkDB*) também está assegurada, e de uma forma bastante simples. Por defeito a própria base de dados já vem preparada para este tipo de escalabilidade e implementa os seus serviços num *cluster* de máquinas, constituído por uma máquina de *query* (responsável por receber os pedidos de acesso a dados) e por várias máquinas *shard* (responsáveis por armazenar os próprios dados).

Deste modo, com a arquitetura proposta anteriormente, o Spidious é capaz de servir um número muito elevado de utilizadores e crescer de forma sustentada. Na verdade, é capaz de crescer até onde as suas próprias ambições o puderem levar, sendo que neste caso o céu (a nuvem) não será o seu limite, mas sim o seu habitat predileto.

5.4 Visualização

A visualização é um dos pontos fulcrais ao qual até o próprio título desta dissertação faz referência. Desse modo esta dissertação, na solução desenvolvida, pretendeu ser o mais inovadora possível no que diz respeito a esta temática. Ao longo de todo o processo foram testadas várias ideias de funcionalidades, sendo que algumas delas resultaram das conclusões retiradas do questionário apresentando anteriormente no [Capítulo 4](#).

A versão final do Spidious é composta essencialmente por uma visualização totalmente orientada ao remetente, muito semelhante às utilizadas numa aplicação de *chat*. Ao longo da mesma destaca-se claramente um painel lateral esquerdo que apresenta a lista das conversas existentes

(individuais e de grupo) e um painel direito que é responsável por apresentar a listagem das *threads* de cada conversação. Está ainda presente uma caixa de resposta rápida, um botão para filtrar as *threads* de trabalho e as de lazer, e um outro botão para resumir o conteúdo de uma *thread*. Apresentar-se-á de seguida cada um dos componentes da interface de forma individual.

O componente de interface visível na [Figura 5.2](#) é o utilizado para listar todas as conversas presentes na interface. Como é possível observar, este componente apresenta o nome de todos os participantes em cada conversa se disponível, caso contrário apresenta o e-mail do participante. Apresenta também a data, em formato relativo, da última mensagem recebida naquela conversa e ainda o número de mensagens por ler em cada uma das já referidas conversas.

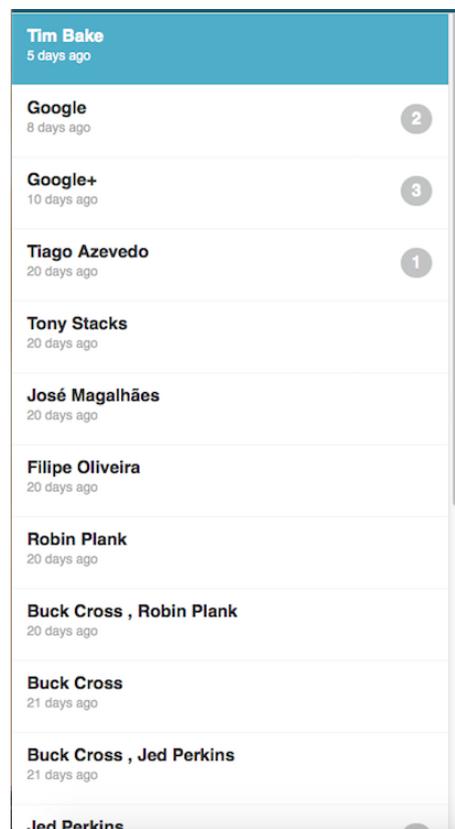


Figura 5.2: Listagem de conversas lateral.

O componente de interface visível na [Figura 5.3](#) é o utilizado para apresentar as diferentes *threads* do e-mail em formato de conversa. No componente são visíveis as datas da última mensagem recebida, o assunto e a classificação de cada *thread*. É de notar que, concetualmente, apenas uma das *threads* se encontra expandida (por defeito será aquela que contenha a mensagem mais recente) sendo que para essa *thread* são ainda visíveis todas as suas mensagens, em forma de balões de conversa, e o remetente de cada uma dessas mensagens.

Spidious: Solução

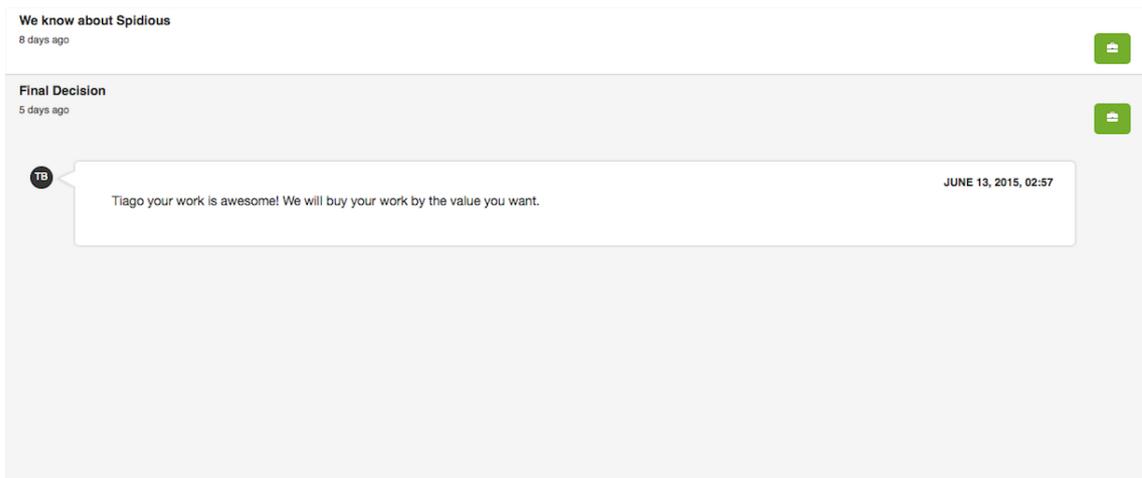


Figura 5.3: Visualização das *threads* em conversa.

O componente de interface visível na [Figura 5.4](#) é utilizado para efetuar uma resposta rápida à *thread* de e-mail expandida naquele momento. Este componente está sempre fixo na parte inferior da interface, para assim possibilitar ao utilizador enviar uma resposta ao mesmo tempo que analisa o conteúdo da própria *thread*. Além do já referido, este componente dá origem a um outro, através do botão *New Subject*, que pode ser consultado na [Figura 5.5](#). Este novo componente tem a função de fornecer ao utilizador a possibilidade de criar um *thread* com um novo assunto, dentro de uma determinada conversação já existente.



Figura 5.4: Caixa de resposta rápida a uma *thread* de e-mail já existente.

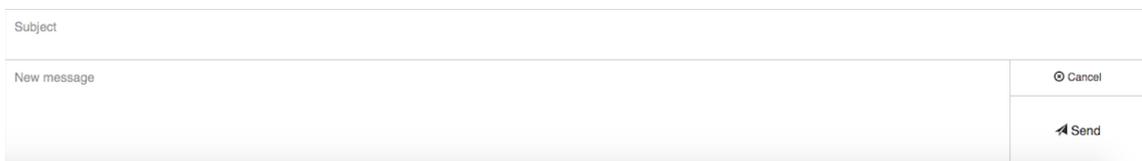


Figura 5.5: Caixa de criação de uma nova *thread*.

O componente de interface visível na [Figura 5.6](#) é um botão que funciona como uma espécie de alternador, que permite mostrar ou esconder o componente de palavras-chave. No momento em que o botão se encontra ativo é apresentado, para todas as *threads*, as palavras-chave de resumo que permitem dessa forma ao utilizador perceber os assuntos tratados ao longo da conversa, antes

mesmo de ler as mensagens que a compõe. O componente que aparece para apresentar as palavras-chave está visível na [Figura 5.7](#). Este componente apresenta sempre as 5 palavras-chave mais relevantes encontradas numa determinada *thread*.



Figura 5.6: Botão para mostrar ou esconder as palavras-chave de resumo para as *threads* de e-mail.

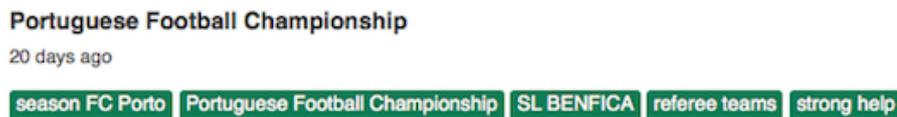


Figura 5.7: Listagem das palavras-chave de resumo de *threads* de e-mail.

O componente de interface visível na [Figura 5.8](#) é utilizado para permitir ao utilizador, em dado momento, a possibilidade de este apenas visualizar/receber mensagens relacionadas com trabalho ou lazer. Caso o utilizador pretenda, poderá também utilizar o filtro *All* e não será aplicado nenhum filtro ao conteúdo da caixa de correio eletrónico.



Figura 5.8: Interface para filtragem de *threads* de e-mail entre trabalho, lazer ou indiferenciado.

O componente de interface visível na [Figura 5.9](#) é um botão que tem como única funcionalidade permitir ao utilizador a criação de uma conversação, com uma lista de participantes ainda não existente.



Figura 5.9: Botão para criação de uma nova conversa.

O componente de interface visível na [Figura 5.10](#) é utilizado nas laterais dos balões que contêm as mensagens referentes a cada conversação. O mesmo tem a funcionalidade de apresentar as informações de um dado remetente de uma mensagem de e-mail, quando o utilizador coloca o rato por cima do elemento. Inicialmente, sem qualquer ação do utilizador, o elemento apresenta as iniciais do nome do remetente se disponível, e em caso contrário, apresenta as iniciais que compõe o seu endereço de e-mail.



Figura 5.10: Interface para mostrar informação dos participantes numa conversa.

Os componente de interface visíveis nas [Figura 5.11](#) e [Figura 5.12](#) assumem a forma de botões que aparecem posicionados na listagem de *threads*, dentro de cada conversa, e para além de mostrarem visualmente a classificação atribuída a determinada *thread*, assumem ainda a funcionalidade de permitir ao utilizador reclassificar aquela mesma *thread*.



Figura 5.11: Identificação para *threads* classificadas como trabalho.



Figura 5.12: Identificação para *threads* classificadas como lazer.

É de referir que a visualização escolhida é apoiada pelas conclusões obtidas no inquérito analisado no [Capítulo 4](#), pese embora que o e-mail não tenha sido pensado para albergar uma visualização deste tipo. Tal facto obrigou a uma grande reestruturação de toda a interface e a vários métodos de adaptação para o utilizador ter uma interação o mais intuitiva possível. No entanto, esta é a mudança de paradigma de visualização no e-mail, com a qual se acredita realmente aumentar a eficiência dos utilizadores, já que é a forma mais natural de tratar a comunicação.

5.5 Inteligência e Aprendizagem

Os algoritmos de inteligência e aprendizagem utilizados no Spidious são importantes no desempenho total do sistema. Também eles compõe o título desta dissertação, e juntamente com a mudança operada na visualização e discutida anteriormente, pretendem fazer parte ativa do aumento de eficiência que o Spidious deseja oferecer aos seus utilizadores. O inquérito levado a cabo e discutido no [Capítulo 4](#) permitiu chegar à conclusão que realmente os utilizadores pretendem um cliente de e-mail que possa ser de alguma forma mais inteligente que os atuais, e permitiu ainda, de acordo com a vontade dos participantes, decidir quais as funcionalidades a implementar. No entanto, o e-mail é uma ferramenta tão crítica na vida das pessoas que foi preciso ter bastante cuidado na delineação das funcionalidades a oferecer nesta temática, uma vez que perder mensagens estava totalmente fora de opção. Desse modo optou-se por oferecer as seguintes funcionalidades:

resumo de *threads* em palavras-chave, classificação de *threads* em trabalho ou lazer e ainda a possibilidade de filtrar a caixa de correio eletrónico a fim de, em dado momento, recebermos apenas o tipo de mensagens que realmente interessam.

O resumo de *threads* foi efetuado sobre a forma de palavras-chave uma vez que, para além de ser o pretendido pelos inquiridos, é o que realmente parece ser a melhor opção para oferecer um resumo, ao qual os utilizadores possam recorrer sempre que necessário, obtendo assim a maior quantidade de informação no menor tempo possível. Para efetuar este resumo foi utilizado o serviço de extração de palavras-chave de seu nome AlchemyAPI ¹⁸, que forneceu um acesso privilegiado ao seu serviço para esta dissertação. A escolha deste serviço deveu-se ao facto de o AlchemyAPI utilizar um algoritmo de extração de palavras-chave, ao mesmo tempo que utiliza outro tipo de informação presente na internet para tentar determinar a importância de cada palavra, ao invés de apenas ter em conta a sua frequência. É ainda de ter em conta que outro fator que pesou na escolha deste serviço prende-se com o facto de o mesmo funcionar em mais de 12 línguas diferentes.

Uma outra funcionalidade, a classificação das *threads* em trabalho e lazer, foi efetuada segundo uma aprendizagem supervisionada. No início o utilizador dá o seu parecer a um conjunto de 20 *threads*, já presente na sua caixa de correio, classificando-as em trabalho ou lazer. O classificador por detrás de todo este processo é um classificador de Bayes que vai sendo treinado em cada classificação manual que o utilizador efetua, utilizando para isso as palavras-chave retiradas de cada *thread* e os e-mails dos participantes presentes nessa mesma conversação. Depois desse passo, o sistema tenta de forma automática classificar as restantes *threads* presentes na caixa de correio eletrónico do utilizador, assim como tentará também classificar uma nova *thread* de e-mail que o utilizador venha a receber no futuro. Caso o classificador cometa algum erro, o utilizador tem total liberdade para retificar a classificação atribuída automaticamente, e dessa forma vai moldando o classificador ao seu próprio perfil de utilização do e-mail.

Por fim, é dada também aos utilizadores a possibilidade de os mesmos ativarem um filtro contínuo que apenas permite visualizar *threads* de trabalho ou lazer. Dessa forma faz-se com que, em dado momento, cada utilizador decida o tipo de mensagens de e-mail que pretende visualizar. Caso o filtro esteja desligado, o utilizador consegue visualizar a classificação atribuída a cada *thread* mas continua a ser importunado simultaneamente por mensagens de e-mail de trabalho e de lazer.

5.6 Resumo da Solução Final

Ao longo desta dissertação foi várias vezes referido que iria ser criada uma nova e moderna interface que levava até ao cliente de e-mail uma diferente abordagem de visualização. Dessa forma, serão agora mostradas algumas imagens que representam determinadas funcionalidades, mas desta feita inseridas de forma integrada na solução final.

¹⁸<http://www.alchemyapi.com/>

Spidious: Solução

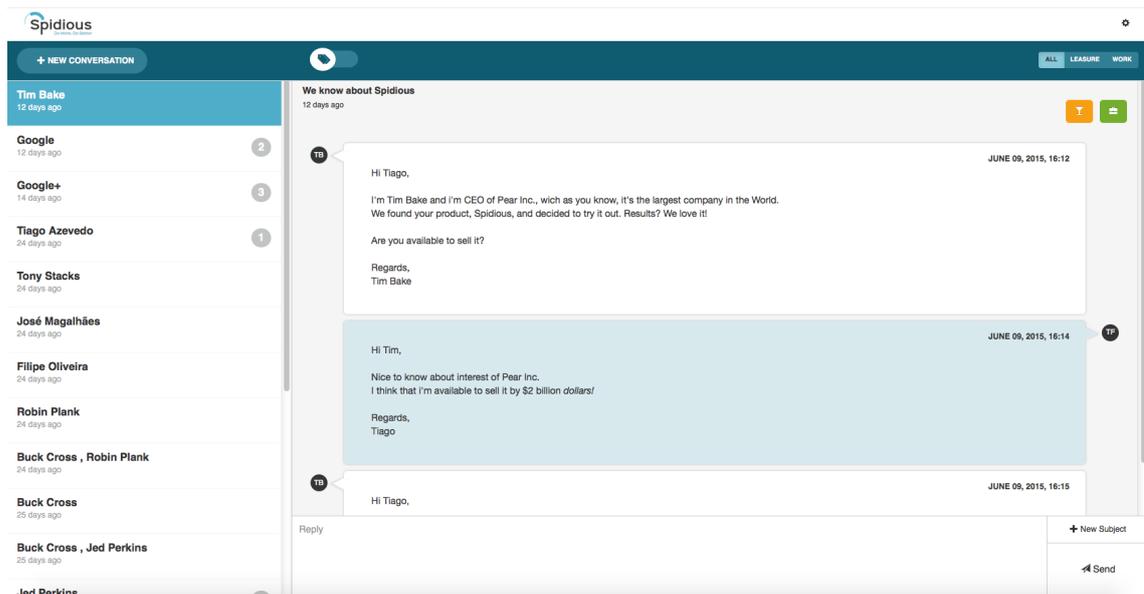


Figura 5.13: Interface da solução na fase de pré-classificação manual.

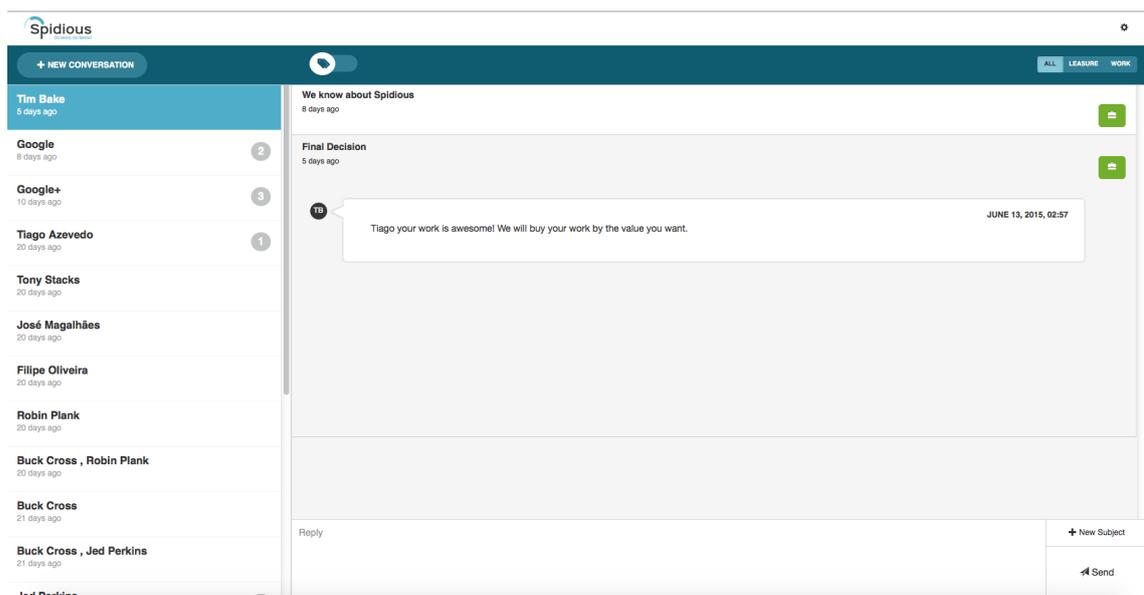


Figura 5.14: Interface da solução após a fase de pré-classificação manual.

Spidious: Solução

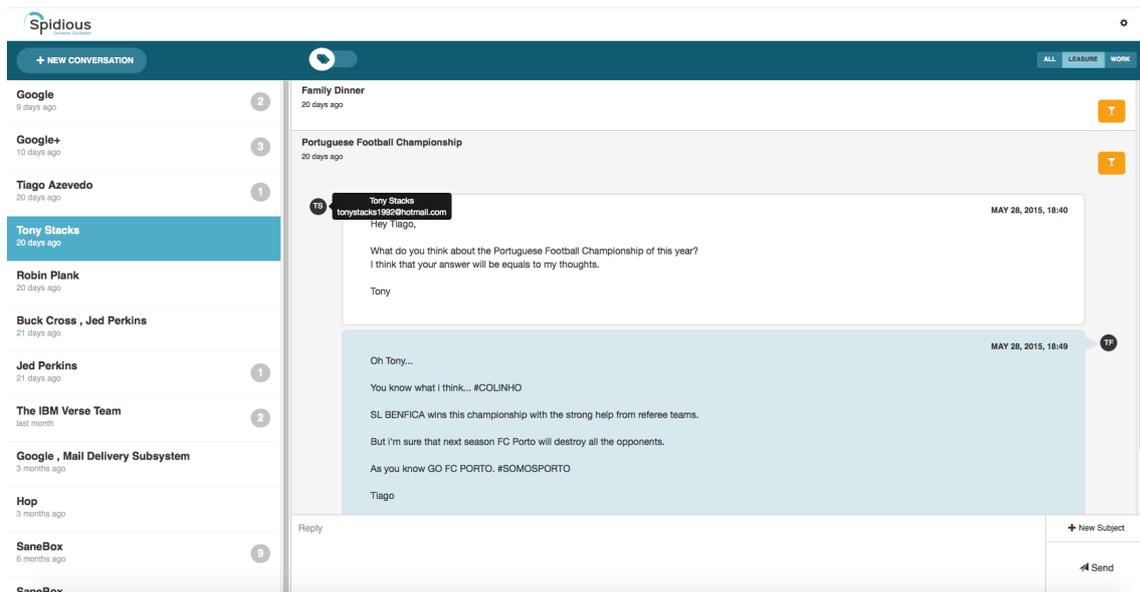


Figura 5.15: Interface da solução com o filtro de lazer ativo mostrando os detalhes de um reme-
tente.

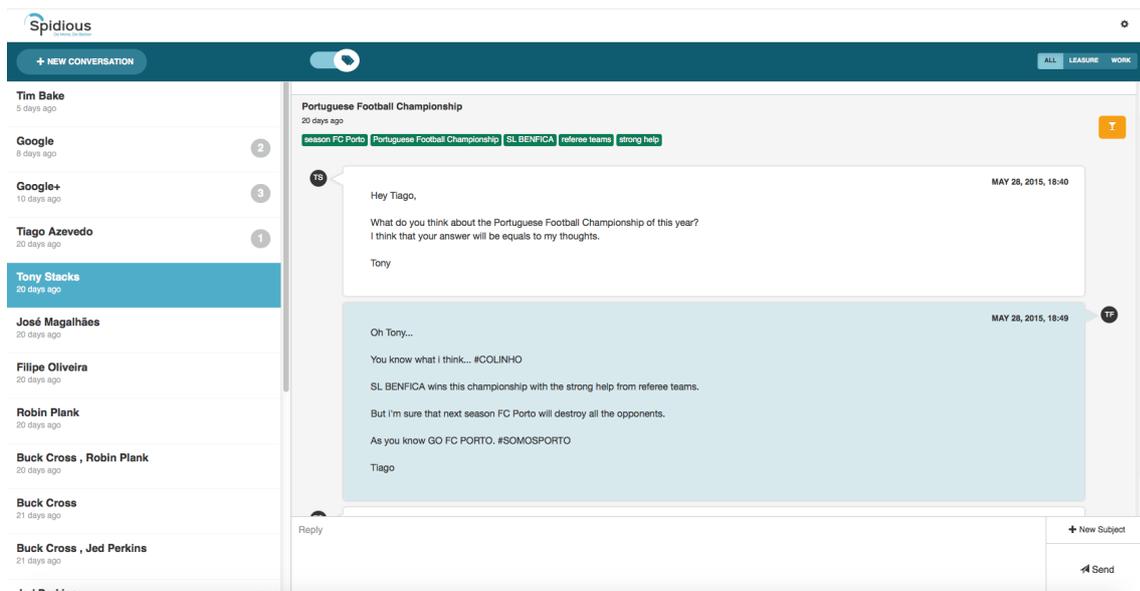


Figura 5.16: Interface da solução com as palavras-chave de resumo ativas.

Spidious: Solução

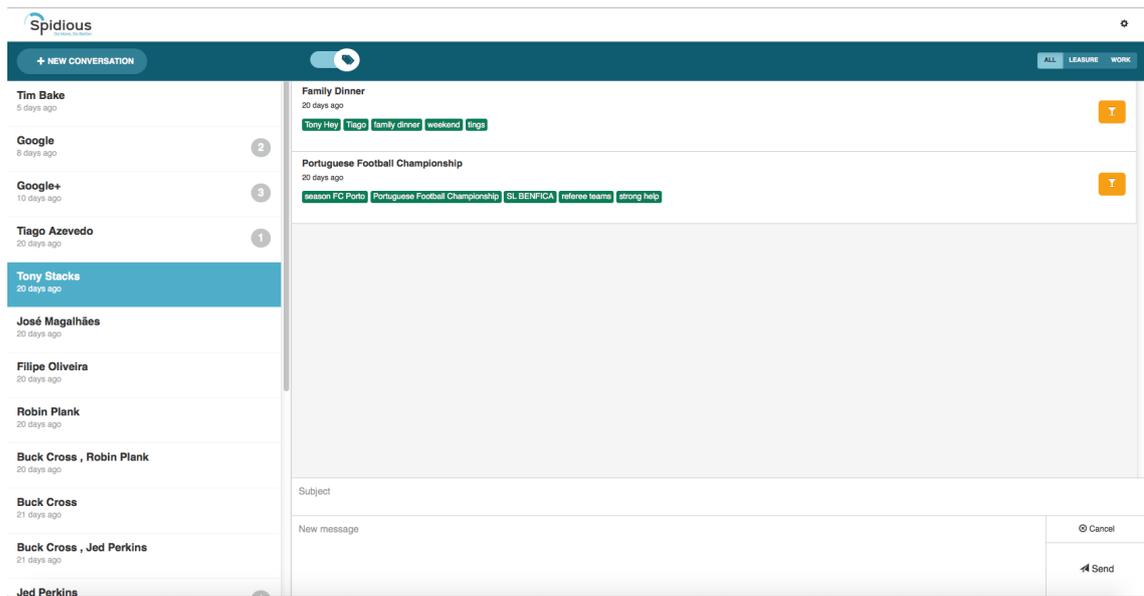


Figura 5.17: Interface da solução aquando da criação de uma nova *thread*.

Spidious: Solução

Capítulo 6

Spidious: Estudo com os utilizadores

Uma solução para um problema com uma dificuldade de resolução tão elevada como aquele que esta dissertação pretende tratar, necessita de alguma forma de ser validado. A validação é uma parte determinante na construção do conhecimento científico uma vez que demonstra a eficácia de teorias e métodos propostos. Contudo, no caso da engenharia de software, torna-se um pouco difícil criar métodos para validar uma determinada técnica ou processo. Devido a este facto, serão utilizadas um conjunto de metodologias que derivam de técnicas descritas no trabalho de Marvin Douglas Dyne, James Landay e Jason Hong [DLH02] para efetuar a validação deste trabalho.

O presente capítulo abordará portanto tudo o que foi feito para tentar aferir a qualidade e o nível de aceitação do Spidious por um conjunto de utilizadores. Apresentar-se-á a metodologia, os resultados e a sua discussão, sendo que logo depois será disponibilizado um pequeno resumo das conclusões obtidas.

6.1 Metodologia

A solução apresentada, e de uma maneira geral toda a dissertação, assume-se como um trabalho bastante ambicioso. O que foi desenvolvido corresponde na sua maior parte a aspetos inovadores, e seguiu o *feedback* recebido no inquérito efetuado à Universidade do Porto. Apesar de se acreditar que o trabalho efetuado é capaz de aumentar a eficiência dos utilizadores no uso do e-mail, é de todo necessário tentar demonstrá-lo. Com tal objetivo em mente foi devidamente desenhado um estudo com utilizadores.

O estudo elaborado compreendia essencialmente três partes diferentes: o preenchimento de um pré-questionário, uma sessão de utilização do Spidious segundo um guião de tarefas e por fim o preenchimento de um pós-questionário. O pré-questionário efetuado serviu essencialmente para garantir que a amostra de sujeitos selecionada era válida. Por seu lado, a sessão de utilização do Spidious, segundo um guião de testes, serviu para dar a cada sujeito do estudo a oportunidade de executar uma série de tarefas no produto, e assim conseguir perceber as suas verdadeiras vantagens. No final, o pós-questionário serviu para obter um conjunto de informações que permitissem

perceber se os sujeitos testados tinham ou não gostado de utilizar o Spidious, para assim poderem apontar os aspetos positivos, os aspetos negativos, e se realmente a solução proposta por esta dissertação se assumia melhor que os seus habituais clientes de e-mail.

O critério de escolha dos utilizadores baseou-se apenas num único fator: o sujeito ser utilizador regular do e-mail. Cada uma das sessões do estudo decorreu no Laboratório de Engenharia de Software da Faculdade de Engenharia da Universidade do Porto e foi efetuado num Macbook Pro 15” (versão 6,2). Em média, cada um dos sujeitos demorou cerca de 50 minutos para completar a sessão de teste completamente. No final de cada um dos testes, todos os dados presentes na solução foram colocados no estado inicial, para que o ambiente do próximo teste fosse exatamente o mesmo.

Resta ainda dizer que as tarefas foram elaboradas tendo em conta um cenário que tentou assemelhar-se o mais possível a um cenário real. Estas mesmas tarefas tentaram avaliar, essencialmente, as novidades introduzidas pelo Spidious deixando assim de fora outras que poderiam processar-se exatamente da mesma forma num cliente de e-mail comum.

6.1.1 Cenário

Para este estudo pretendia-se que cada sujeito se visse integrado num contexto o mais real possível. Desse modo foi criado um cenário inicial que os sujeitos tiveram de considerar:

“Considere que se chama Tiago Costa e é um estudante finalista no Mestrado Integrado em Engenharia Informática e Computação da Faculdade de Engenharia da Universidade do Porto. No âmbito da sua dissertação, o Tiago desenvolveu um cliente de e-mail web, de seu nome Spidious, que muda o paradigma comum aos clientes de e-mail atuais. De momento, o Tiago encontra-se a utilizar o Spidious no seu dia-a-dia e de entre as suas diversas conversações, uma delas é com Tim Bake o CEO da Pear Inc., a maior empresa de software do mundo, que está a negociar a compra do Spidious com o Tiago.”

É de referir ainda que a conta de e-mail envolvida neste cenário tinha cerca de 200 mensagens de e-mail, na sua maioria reais, sendo que os restantes foram criados para serem usados neste estudo. Para levar a cabo esta adição de mensagens extra foram criadas contas de e-mail que representavam remetentes fictícios, com perfis bem definidos os quais influenciavam o tipo de conversas em que estes estavam envolvidos. Os referidos remetentes fictícios e os seus perfis serão agora apresentados:

Jed Perkins Este remetente assumiu o perfil de amigo do Tiago Costa.

Tony Stacks Este remetente assumiu o perfil de familiar do Tiago Costa.

Buck Cross Este remetente assumiu o perfil de colega de trabalho e amigo do Tiago Costa.

Robin Plank Este remetente assumiu o perfil de colega de trabalho do Tiago Costa.

Tim Bake Este remetente assumiu o perfil de CEO da Pear Inc., empresa já descrita no cenário (Subsecção 6.1.1), e portanto representava também ele um contacto relacionado com trabalho.

6.1.2 Tarefas

Baseados no cenário referido anteriormente na Subsecção 6.1.1 os sujeitos do estudo com utilizadores realizaram um conjunto de tarefas. Ao todo foram definidas sete tarefas, descritas no guião de tarefas da Secção B.3, que foram pensadas para permitir que os participantes fossem capazes de experimentar todas as novas funcionalidades oferecidas pelo Spidious. Seguiremos agora para a análise individual de cada uma dessas tarefas:

Tarefa 1 A primeira tarefa consistia em pedir aos sujeitos para classificarem vinte *threads* de e-mail diferentes, em trabalho ou lazer, ao longo da lista de conversações disponível. Depois disso era comunicado ao utilizador que o Spidious iria classificar todas as outras *threads* que permaneciam sem classificação. Esta tarefa teve como objetivo principal permitir ao utilizador testar o sistema de classificação da solução, e dessa forma poder avaliá-lo.

Tarefa 2 A segunda tarefa consistia em pedir aos sujeitos que encontrassem uma *thread*, na conversa com o Tony Stacks, que falava sobre o campeonato português de futebol. Esta tarefa teve como objetivo permitir aos participantes testarem a orientação ao remetente do Spidious, a fim de conseguirem perceber se a mesma era, ou não, vantajosa.

Tarefa 3 Na terceira tarefa era pedido aos sujeitos que respondessem à última mensagem deixada pelo Tony Stacks na *thread* sobre o campeonato português de futebol. Esta tarefa tinha como objetivo dar aos participantes a oportunidade de experimentarem a interface de resposta rápida a uma *thread*, implementada no Spidious.

Tarefa 4 Na seguinte tarefa (Tarefa 4) foi pedido aos sujeitos que ativassem a funcionalidade de resumo de *threads* com recurso a palavras-chave. Depois disso era ainda pedido para analisarem as palavras-chave geradas. Esta tarefa tinha como objetivo fazer os utilizadores experimentarem a funcionalidade de resumo de *threads* do Spidious, para dessa forma poderem avaliá-la.

Tarefa 5 A quinta tarefa consistia em pedir aos participantes que, na conversa com o Tony Stacks, fosse criada uma nova *thread* com um assunto diferente dos já existentes. Esta tarefa tinha como objetivo perceber se os utilizadores gostavam da interface inovadora que foi desenvolvida para esta funcionalidade.

Tarefa 6 A sexta tarefa consistia em pedir aos utilizadores para efetuarem um clique no botão com a funcionalidade de criar uma nova conversa. O seu principal objetivo era perceber se a localização do botão era a ideal.

Tarefa 7 Na última tarefa (Tarefa 7) foi pedido aos sujeitos que aplicassem um filtro à caixa de correio eletrónico, para que dessa forma apenas lhes fossem mostradas as mensagens classificadas como lazer. O principal objetivo desta tarefa foi perceber se a interface para efetuar a filtragem era intuitiva, e paralelamente permitir ao utilizador avaliá-la e perceber claramente a sua utilidade.

6.2 Resultados

O estudo realizado contou com vinte utilizadores, e este não foi de todo um número escolhido ao acaso. Segundo o artigo do Nielson Norman Group ¹, para um estudo com utilizadores poder apresentar resultados estatisticamente significantes, o número ideal de utilizadores de teste fixa-se em vinte.

Os resultados provenientes deste estudo serviram por isso para validar a solução apresentada nesta dissertação, e determinar o quão melhor a mesma é em relação aos clientes de e-mail habituais dos participantes. Os referidos resultados permitiram ainda dar resposta às questões de pesquisa descritas no início desta dissertação na [Secção 1.3](#).

6.2.1 Pré-Questionário

Na [Secção 6.1](#) é referido que o pré-questionário foi efetuado com o objetivo concreto de validar a amostra de sujeitos de teste escolhidos. Como se pode verificar na [Secção B.1](#) do [Apêndice B](#) o pré-questionário é praticamente idêntico ao questionário efetuado no inquérito à Universidade do Porto ([Apêndice A](#)).

Depois de analisados os resultados chegou-se à conclusão que os mesmos não apresentavam um desvio significativo dos encontrados no inquérito descrito no [Capítulo 4](#). Desse modo concluiu-se que a amostra utilizada neste estudo estava de acordo com a população encontrada no inquérito realizado previamente à Universidade do Porto. Os dados recolhidos neste pré-inquérito poderão ser consultados em pormenor na [Secção B.2](#).

6.2.2 Guião de Tarefas

O guião de tarefas desenvolvido, além de apresentar o cenário para o teste a efetuar e a consequente lista de tarefas, apresentava também um conjunto de questões no final do enunciado de cada tarefa. A maioria dos resultados das respostas a estas questões apenas têm utilidade para ser possível analisar a validade de cada sessão de testes. No entanto existem duas perguntas que foram criadas, no guião de tarefas, com a finalidade de poder ser analisada a percentagem de acerto do classificador automático. Essas mesmas questões terão os resultados apresentados, já de seguida, nesta secção. Os restantes resultados podem ser consultados na íntegra na [Secção B.4](#) do [Apêndice B](#).

¹<http://www.nngroup.com/articles/how-many-test-users/>

Spidious: Estudo com os utilizadores

Assim sendo falar-se-á agora do método já referido para aferir o acerto do classificador automático. Na primeira tarefa presente no guião é pedido ao utilizador para efetuar a classificação manual de vinte *threads* de e-mail para que o Spidious possa, de alguma forma, começar a traçar as preferências para esse inquirido. Depois disso é perguntado ao participante para, segundo a sua perceção pessoal, atribuir e registar a sua classificação para a seguinte mensagem de e-mail: “*Tiago your work is awesome! We’ll buy your work by the value you want*”. O gráfico dos resultados obtidos para esta questão pode ser consultado na [Figura 6.1](#).

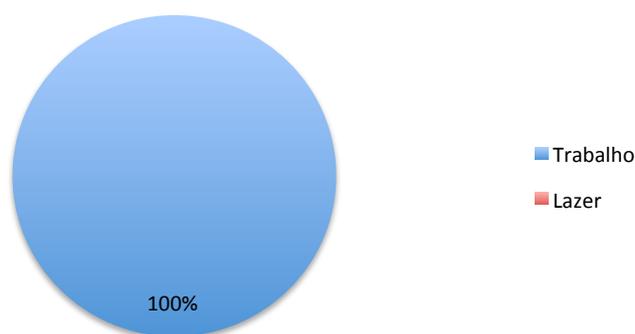


Figura 6.1: Opinião sobre a classificação a atribuir à seguinte mensagem de e-mail: “*Tiago your work is awesome! We’ll buy your work by the value you want*”.

Depois de anotada a resposta do utilizador, a mensagem por ele classificada anteriormente era inserida numa nova *thread* no Spidious, que a tentava classificar de forma automática. Após a inserção era pedido ao inquirido para verificar qual a classificação automática atribuída pelo Spidious à mensagem, e pedido para o mesmo a registar também. O gráfico dos resultados da classificação automática pode ser consultado na [Figura 6.2](#).

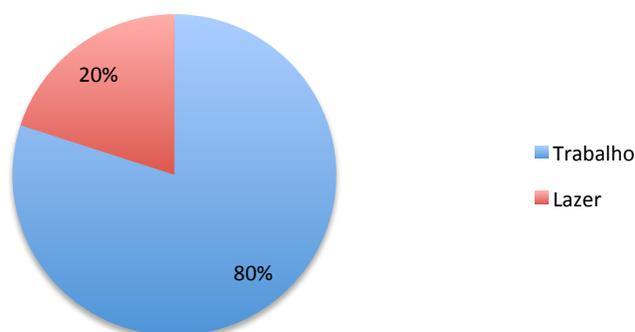


Figura 6.2: Classificação automática atribuída pelo Spidious à mensagem anterior (A mensagem anterior faz parte de uma nova *thread* recebida na conversação com Tim Bake).

A partir deste último gráfico apresentado é possível constatar que a percentagem de acerto do classificador automático oferecido pelo Spidious, neste teste com utilizadores, foi de 80%.

6.2.3 Pós-Questionário

A presente secção analisará os resultados do pós-questionário efetuado logo depois de cada sujeito ter executado a sua sessão de utilização no Spidious. Estes serão os resultados que usaremos para retirar as conclusões finais do trabalho e determinar as suas vantagens e pontos fracos, caso existam.

A análise foi dividida segundo as categorias do questionário: tarefas, funcionalidades e considerações finais. A categoria das tarefas serviu essencialmente para perceber o nível de dificuldade que cada sujeito sentiu ao executar uma determinada tarefa. Já a categoria das funcionalidades serviu para reunir a opinião dos sujeitos em relação às funcionalidades desenvolvidas no Spidious, quer a nível da sua utilidade, quer a nível da qualidade do seu resultado final. Por fim, a categoria das considerações finais serve para perceber quão melhor os sujeitos consideram o Spidious, face ao seu cliente de e-mail habitual, e se estariam dispostos a utilizá-lo no futuro.

6.2.3.1 Tarefas

A primeira categoria deste questionário permite perceber o nível de dificuldade que os participantes sentiram ao executar cada uma das tarefas pedidas. Esta foi uma categoria bastante importante, uma vez que a partir da mesma é possível ter uma primeira ideia das funcionalidades que podem ser melhoradas a nível de interação. Tal observação é inferida a partir do facto de que componentes de interação bem definidos apresentarão valores elevados para as dificuldades de conclusão “Muito Fácil” e “Fácil”.

Os primeiros três gráficos analisados nesta categoria, resumem então a informação da dificuldade de conclusão da tarefa 1 (Figura 6.3), da dificuldade de conclusão da tarefa 2 (Figura 6.4) e da dificuldade de conclusão da tarefa 3 (Figura 6.5). A partir destes gráficos é possível constatar que 65% e 25% dos participantes sentiram que a tarefa 1 teve um nível de dificuldade de “Muito Fácil” e “Fácil” respetivamente, que 80% dos participantes sentiram que a tarefa 2 apresentou um nível de dificuldade “Muito Fácil” e ainda por fim que 90% sentiu que a tarefa 3 teve um nível de dificuldade “Muito Fácil”.

Spidious: Estudo com os utilizadores

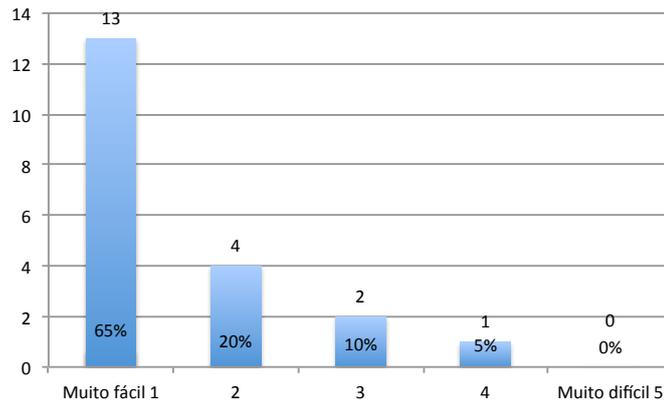


Figura 6.3: Nível de dificuldade sentido na execução da tarefa 1 - classificação de algumas *threads* de e-mail. $\bar{x} \approx 1.55$; $\sigma \approx 0.89$

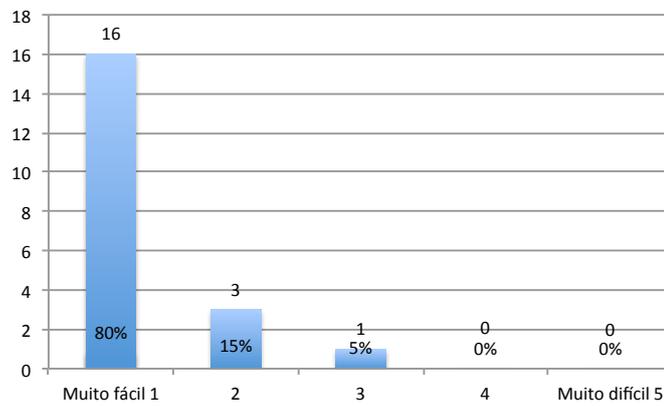


Figura 6.4: Nível de dificuldade sentido na execução da tarefa 2 - pesquisa de uma *thread*. $\bar{x} \approx 1.25$; $\sigma \approx 0.55$

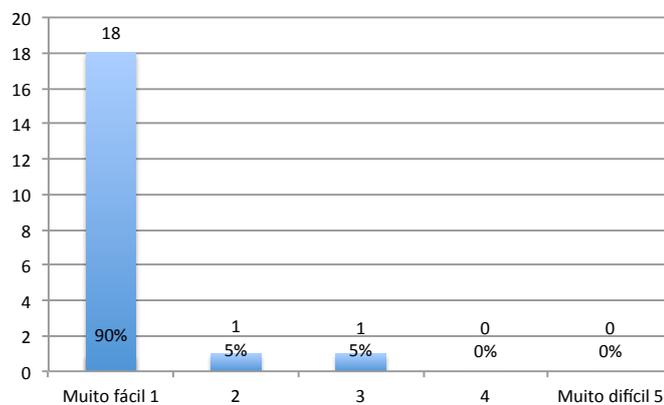


Figura 6.5: Nível de dificuldade sentido na execução da tarefa 3 - resposta a uma mensagem de e-mail. $\bar{x} \approx 1.15$; $\sigma \approx 0.49$

Os seguintes gráficos a serem analisados resumem a informação da dificuldade de conclusão da tarefa 4 (Figura 6.6), da tarefa 5 (Figura 6.7), da tarefa 6 (Figura 6.8) e da tarefa 7 (Figura 6.9). Com recurso a estes gráficos é possível constatar que apenas 40% dos participantes sentiram que a tarefa 4 teve um nível de dificuldade de “Muito Fácil”, que 90% dos participantes sentiram que a tarefa 5 apresentou um nível de dificuldade “Muito Fácil”, que 95% dos participantes sentiram que a tarefa 6 apresentou um nível de dificuldade “Muito Fácil” e ainda por fim que 85% sentiu que a tarefa 7 teve um nível de dificuldade “Muito Fácil”.

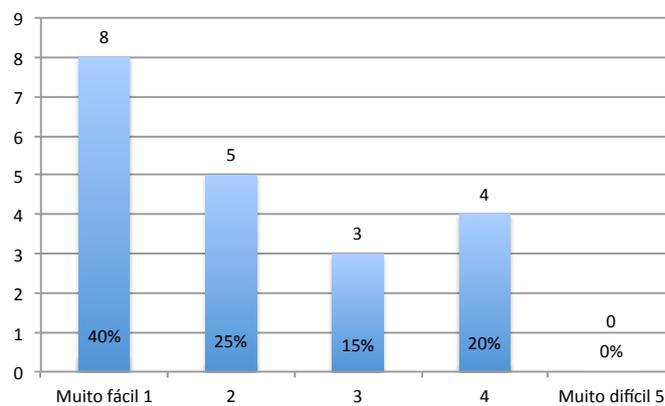


Figura 6.6: Nível de dificuldade sentido na execução da tarefa 4 - obtenção das palavras-chave para resumo de *threads*. $\bar{x} \approx 2.15$; $\sigma \approx 1.18$

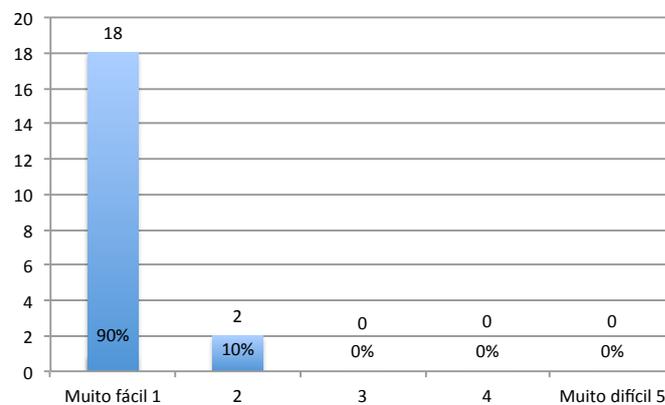


Figura 6.7: Nível de dificuldade sentido na execução da tarefa 5 - criação de uma nova *thread*. $\bar{x} \approx 1.1$; $\sigma \approx 0.31$

Spidious: Estudo com os utilizadores

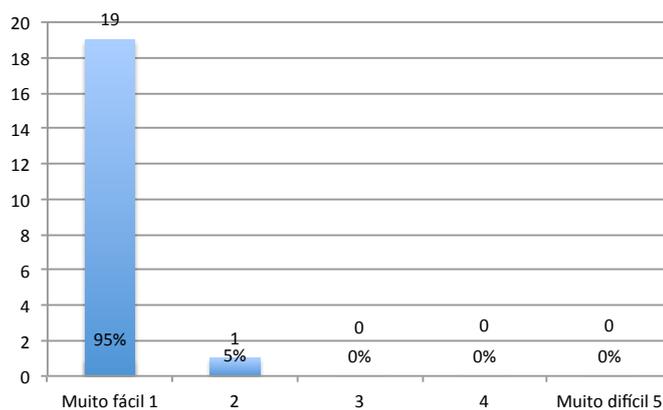


Figura 6.8: Nível de dificuldade sentido na execução da tarefa 6 - utilização do botão de criação de uma nova conversa. $\bar{x} \approx 1.05$; $\sigma \approx 0.22$

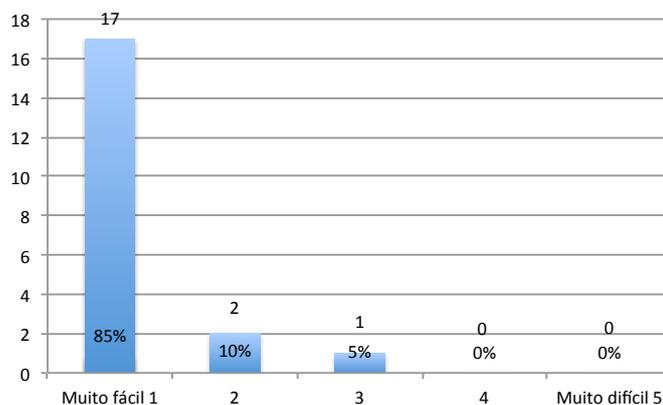


Figura 6.9: Nível de dificuldade sentido na execução da tarefa 7 - aplicação de filtros à caixa de correio eletrónico. $\bar{x} \approx 1.2$; $\sigma \approx 0.52$

6.2.3.2 Funcionalidades

A próxima categoria do pós-questionário a ser analisada é a das funcionalidades, sendo que esta apresenta um grau de importância elevado. Esta categoria do questionário teve como objetivo adquirir informação sobre a opinião dos utilizadores acerca da utilidade e da qualidade do resultado final da solução. A categoria possui bastantes gráficos que têm de ser analisados, pelo que os mesmos serão analisados em grupos. O grupo que inicia esta análise é composto por quatro gráficos.

O primeiro gráfico (Figura 6.10) apresenta informação acerca da utilidade da visualização orientada à conversa presente no Spidious. É possível observar que 50% dos inquiridos afirmam que a referida funcionalidade é “Muito Útil” e que 40% afirma que é “Útil”. O segundo gráfico (Figura 6.11) apresenta informação acerca da utilidade da extração de palavras-chave, para resumo de *threads*, apresentada no Spidious. É possível constatar que 45% dos inquiridos afirma que a

Spidious: Estudo com os utilizadores

referida funcionalidade é “Muito Útil” e que 55% afirma que é “Útil”. O terceiro gráfico (Figura 6.12) apresenta informação acerca da utilidade da classificação de *threads*, em trabalho ou lazer, apresentada no Spidious. É possível perceber que 45% dos inquiridos afirma que a referida funcionalidade é “Muito Útil” e que também 45% afirma que é “Útil”. Por último, o quarto gráfico (Figura 6.13) apresenta informação acerca da utilidade da filtragem por mensagens indiferenciadas, de trabalho ou lazer, apresentada no Spidious. É possível observar que 65% dos inquiridos afirma que a referida funcionalidade é “Muito Útil” e que 30% afirma que é “Útil”.

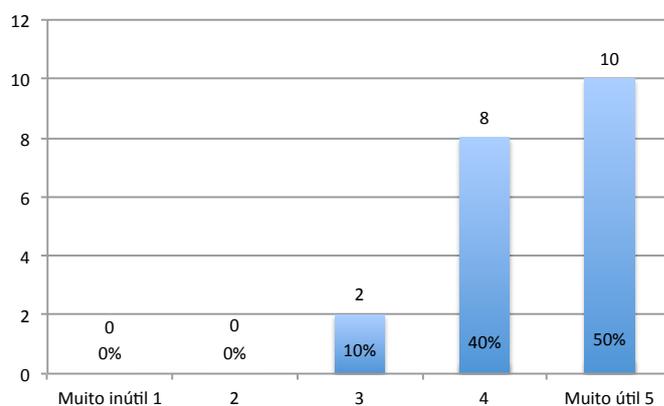


Figura 6.10: Opinião sobre a utilidade da visualização orientada à conversa apresentada no Spidious. $\bar{x} \approx 4.4$; $\sigma \approx 0.68$

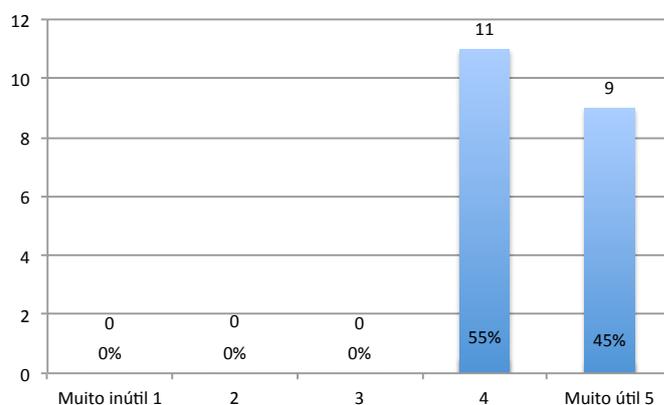


Figura 6.11: Opinião sobre a utilidade da extração de palavras-chave, para resumo de *threads*, apresentada no Spidious. $\bar{x} \approx 4.45$; $\sigma \approx 0.51$

Spidious: Estudo com os utilizadores

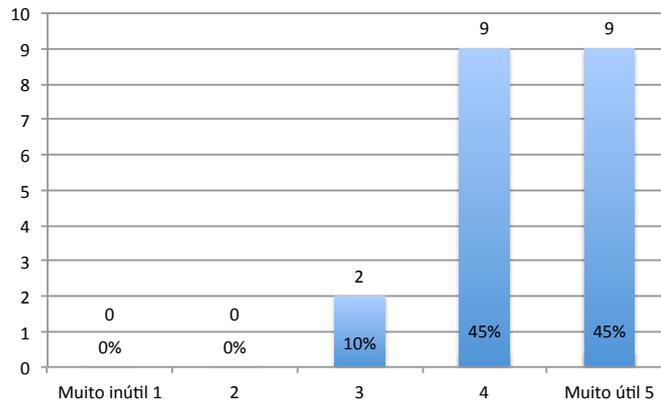


Figura 6.12: Opinião sobre a utilidade da classificação de *threads*, em trabalho ou lazer, apresentada no Spidious. $\bar{x} \approx 4.45$; $\sigma \approx 0.67$

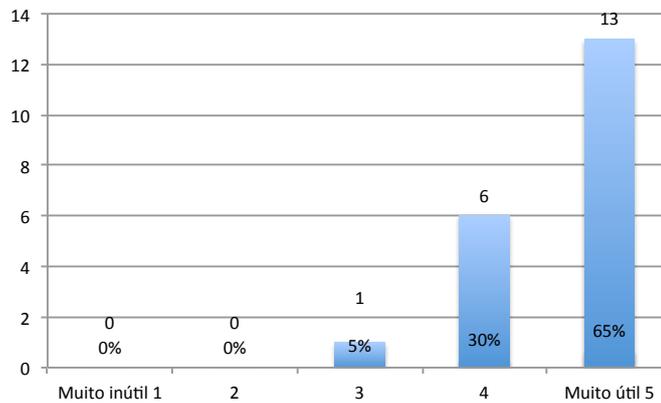


Figura 6.13: Opinião sobre a utilidade da filtragem por mensagens indiferenciadas, de trabalho ou de lazer, apresentada no Spidious. $\bar{x} \approx 4.6$; $\sigma \approx 0.60$

O grupo que se segue nesta análise é também composto por quatro gráficos. O primeiro gráfico (Figura 6.14) apresenta informação sobre o resultado final da visualização orientada à conversa apresentada no Spidious. É possível observar que 65% dos inquiridos afirma que o resultado final é “Muito Bom” e que 35% afirma que é “Bom”. O segundo gráfico (Figura 6.15) apresenta informação acerca do resultado final da extração de palavras-chave, para resumo de *threads*, apresentada no Spidious. É possível constatar que 40% dos inquiridos afirma que o resultado final é “Muito Bom” e que 50% afirma que é “Bom”. O terceiro gráfico (Figura 6.16) apresenta informação acerca do resultado final da classificação de *threads*, em trabalho ou lazer, apresentada no Spidious. É possível perceber que 30% dos inquiridos afirma que o resultado final é “Muito Bom” e que também 60% afirma que é “Bom”. Por último, o quarto gráfico (Figura 6.17) apresenta informação acerca do resultado final da filtragem por mensagens indiferenciadas, de trabalho ou lazer, apresentada no Spidious. É possível observar que 45% dos inquiridos afirma que o resultado final é “Muito Bom” e que 50% afirma que é “Bom”.

Spidious: Estudo com os utilizadores

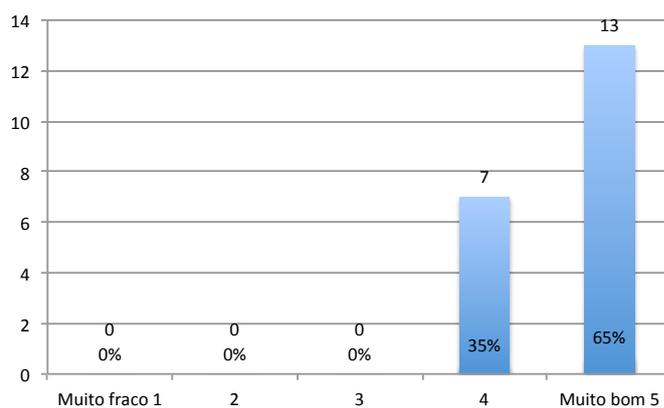


Figura 6.14: Opinião sobre o resultado final da visualização orientada à conversa apresentada no Spidious. $\bar{x} \approx 4.65$; $\sigma \approx 0.49$

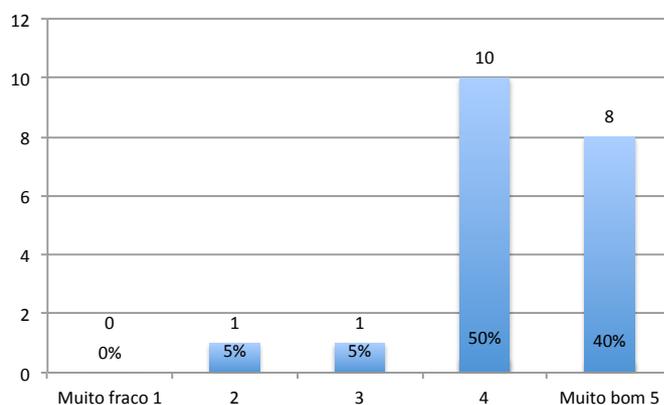


Figura 6.15: Opinião sobre o resultado final da extração de palavras-chave, para resumo de *threads*, apresentada no Spidious. $\bar{x} \approx 4.25$; $\sigma \approx 0.79$

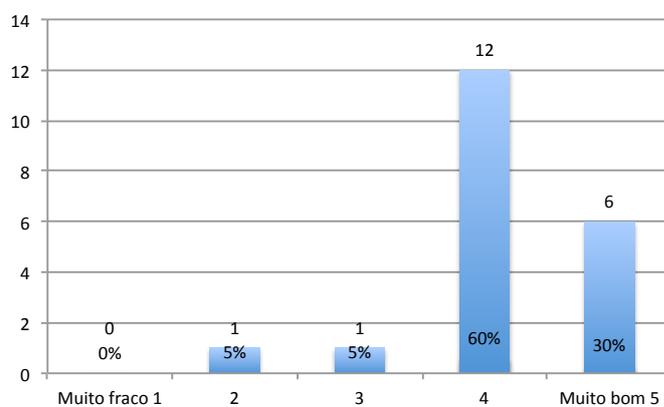


Figura 6.16: Opinião sobre o resultado final da classificação de *threads*, em trabalho ou lazer, apresentada no Spidious. $\bar{x} \approx 4.15$; $\sigma \approx 0.75$

Spidious: Estudo com os utilizadores

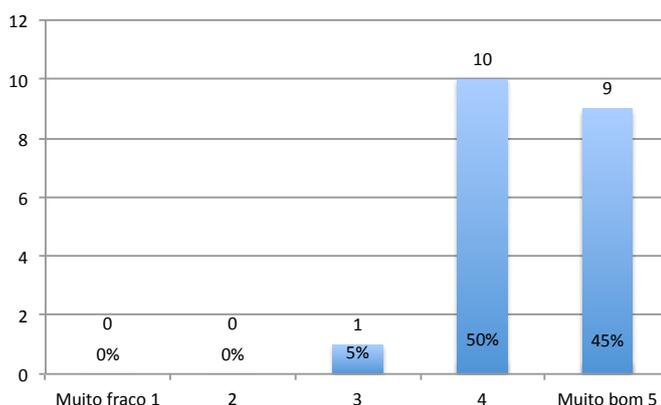


Figura 6.17: Opinião sobre o resultado final da filtragem por mensagens indiferenciadas, de trabalho ou de lazer apresentada no Spidious. $\bar{x} \approx 4.45$; $\sigma \approx 0.60$

Para terminar, o último gráfico (Figura 6.18) a ser analisado trata a opinião dos utilizadores em relação a quão fácil foi o uso de toda a interface do Spidious, resultante da mudança de paradigma no e-mail operada pelo mesmo. Nesse sentido é observável que 45% dos participantes considera que a interface da solução apresenta uma usabilidade “Muito Fácil” enquanto que 40% considera que a usabilidade é “Fácil”.

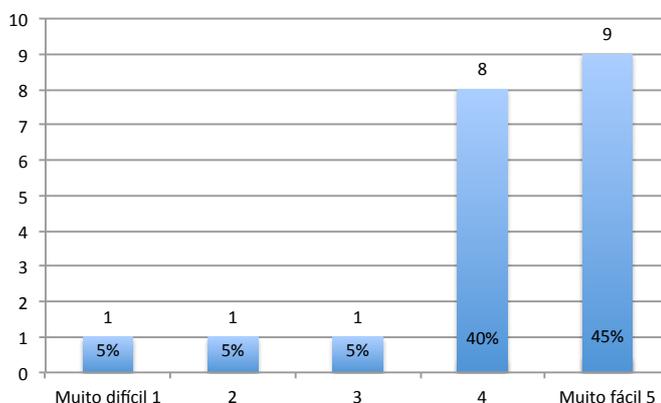


Figura 6.18: Opinião sobre quão fácil é o uso de toda a interface do Spidious, resultante da mudança de paradigma no e-mail operada pelo mesmo. $\bar{x} \approx 4.15$; $\sigma \approx 1.09$

6.2.3.3 Considerações Finais

A última categoria do pós-questionário serviu em primeiro lugar para registar a opinião final dos participantes em relação ao Spidious, perceber o que gostaram e o que não gostaram tanto. Em segundo lugar serviu também para reunir informação que diretamente pudesse responder às questões de pesquisa às quais esta dissertação se propõe a responder.

Desse modo analisar-se-ão de seguida os três primeiros gráficos de resultados desta categoria. O primeiro deles (Figura 6.19) expõe informação acerca da opinião dos participantes sobre o

Spidious: Estudo com os utilizadores

aumento de eficiência de utilização que os novos métodos de visualização do Spidious conferem ao e-mail, face aos clientes habituais dos utilizadores. É perfeitamente observável no gráfico que a totalidade dos participantes no estudo respondeu afirmativamente.

O segundo gráfico (Figura 6.20) retrata informação acerca da opinião dos participantes sobre o aumento de eficiência de utilização que os algoritmos de aprendizagem do Spidious conferem ao e-mail, face aos clientes habituais dos utilizadores. Novamente, constata-se que 100% dos inquiridos respondeu afirmativamente.

Por fim, o terceiro e último gráfico deste primeiro conjunto (Figura 6.21) apresenta informação acerca da opinião dos inquiridos sobre o aumento de eficiência de utilização que a conjugação de uma nova visualização e de técnicas de aprendizagem num cliente web (Spidious) conferem ao e-mail, face aos clientes habituais dos utilizadores. A totalidade dos participantes no estudo respondeu afirmativamente.

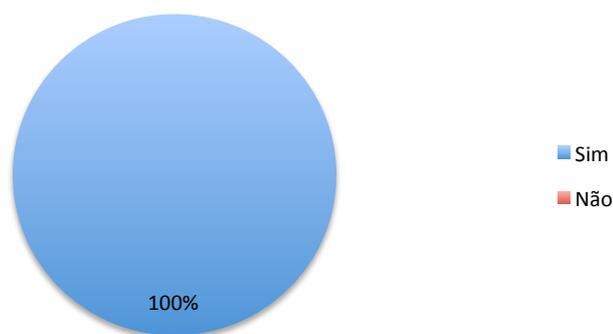


Figura 6.19: Opinião sobre o aumento de eficiência de utilização que os novos métodos de visualização do Spidious conferem ao e-mail, face aos clientes habituais dos utilizadores.

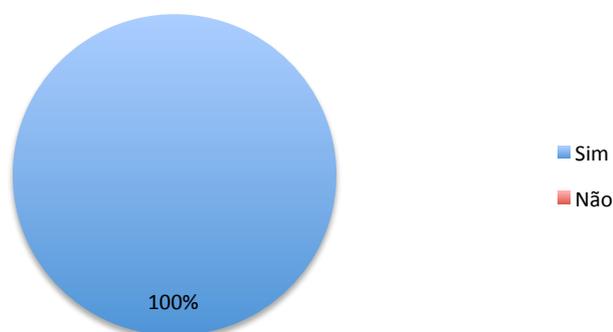


Figura 6.20: Opinião sobre o aumento de eficiência de utilização que os algoritmos de aprendizagem do Spidious conferem ao e-mail, face aos clientes habituais dos utilizadores.

Spidious: Estudo com os utilizadores

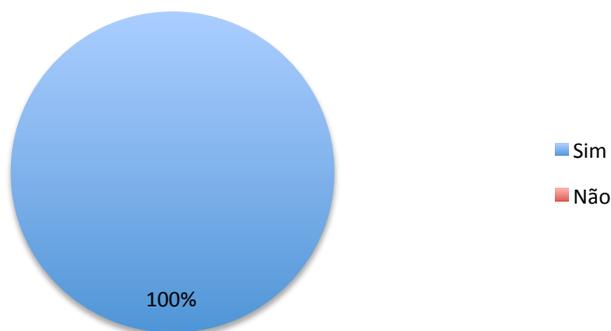


Figura 6.21: Opinião sobre o aumento de eficiência de utilização que a conjugação de uma nova visualização e de técnicas de aprendizagem num cliente de e-mail web (Spidious) poderiam conferir, face aos clientes habituais dos utilizadores.

De forma a continuar a análise aos resultados obtidos nesta categoria analisaremos agora três gráficos. O primeiro deles (Figura 6.22) analisa a opinião sobre quão melhor os utilizadores acharam a experiência de utilização proporcionada pelo Spidious, em comparação à oferecida pelos seus clientes de e-mail habituais. O segundo (Figura 6.23) fornece informação sobre a vontade de os inquiridos virem a utilizar a solução no futuro. Por último, o terceiro (Figura 6.24) disponibiliza informação acerca da disponibilidade dos inquiridos pagarem pela solução. Recorrendo aos gráficos é possível constatar que pelo menos 90% dos participantes considera que o Spidious oferece pelo menos uma “Melhor” experiência de utilização que a do seu cliente de e-mail habitual. É possível observar também que a esmagadora maioria (95%) estava disposto a utilizá-lo no futuro caso o mesmo chegasse a produção, e por fim, 85% dos inquiridos afirma que não estaria disposto a pagar qualquer tipo de mensalidade para utilizar o produto.

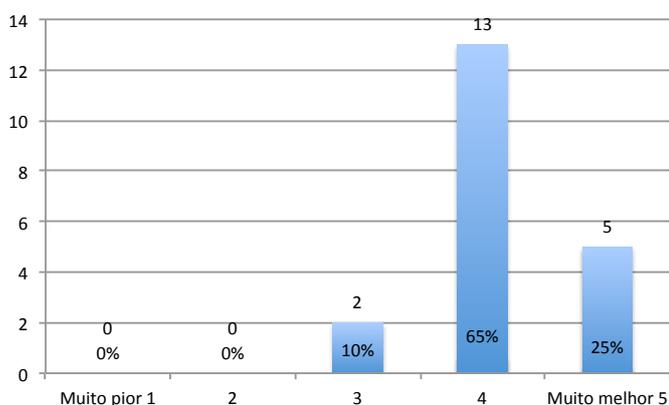


Figura 6.22: Opinião sobre quão melhor os utilizadores acharam a experiência de utilização proporcionada pelo Spidious, em comparação com a fornecida pelos seus clientes de e-mail habituais. $\bar{x} \approx 4.15$; $\sigma \approx 0.59$

Spidious: Estudo com os utilizadores

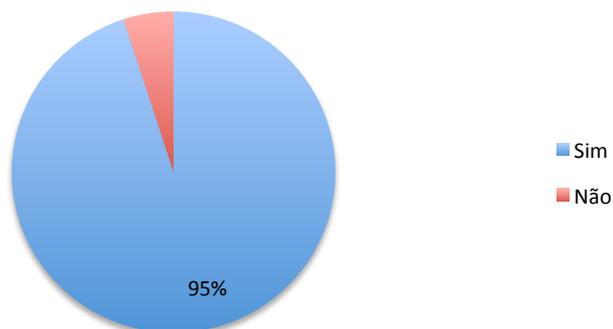


Figura 6.23: Opinião dos utilizadores em relação a uma futura utilização do Spidious, caso o mesmo chegasse à fase de produção.

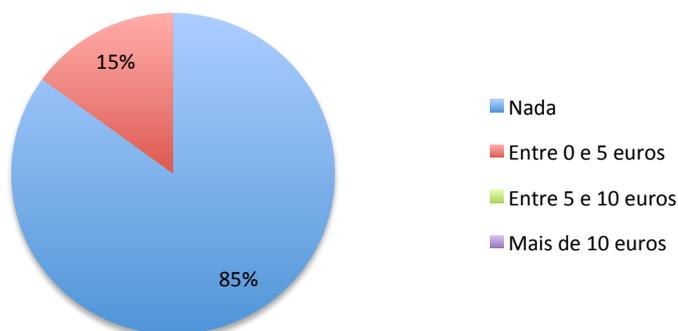


Figura 6.24: Opinião dos utilizadores em relação à disponibilidade de, no futuro, virem a pagar uma subscrição mensal para utilizarem o Spidious.

O questionário termina com duas questões de resposta aberta, onde é perguntado aos participantes do estudo para apontarem aquilo que mais gostaram, e também o que menos gostaram, do Spidious. Os resultados obtidos encontram-se resumidos, e agrupados, na [Tabela 6.1](#) apresentada de seguida.

Spidious: Estudo com os utilizadores

Aspetos Positivos do Spidious	Aspetos Negativos do Spidious
Filosofia de interface diferente e bem conseguida.	Inexistência de teclas de atalho.
Classificação das <i>threads</i> em trabalho e lazer.	Não existência da funcionalidade de gestão de tarefas.
Aprendizagem orientada ao utilizador.	Ausência de uma barra de pesquisa.
Palavras-chave de resumo de <i>threads</i> .	O aspeto visual das palavras-chave de resumo poderia ser ainda mais destacado.
Resposta a uma mensagem de e-mail muito rápida e fácil.	Não permitir a criação de mais filtros.
Possibilidade de filtrar as mensagens de e-mail por categorias.	A última <i>thread</i> de cada conversação aparece, por defeito, expandida.
Interface simples, limpa, organizada, intuitiva e mais baseada em imagens que em texto.	O botão que permite ligar as palavras-chave de resumo de <i>threads</i> pode ser melhorado, uma vez que apenas o ícone é insuficiente.
Visualização orientada ao remetente, permitindo que a divisão principal da caixa de correio eletrónico seja feita com recurso aos contactos.	As palavras-chave, correspondentes à <i>thread</i> que contém a mensagem mais recente de cada contacto, poderiam já aparecer na lista de contactos.

Tabela 6.1: Lista dos aspetos positivos e negativos, apontados pelos participantes do inquérito, do Spidious.

6.3 Discussão dos Resultados

A discussão de resultados que se segue apresentará o conteúdo mais importante para retirar as conclusões de todo o trabalho e até mesmo poder responder às questões de pesquisa já levantadas. É perceptível, por essa razão, que esta secção contém uma das partes mais relevantes desta dissertação.

Como já referido anteriormente, a primeira parte do pós-questionário reuniu informação sobre a dificuldade de cada tarefa executada no teste com utilizadores. Uma vez que cada tarefa foi pensada para explorar cada um desses componentes de interface, a referida parte do pós-questionário é ideal para perceber quais os componentes de interface que necessitavam de ser mais trabalhados. Depois de apresentados os resultados constata-se que a tarefa que nos seus resultados apresentou maior nível de dificuldade foi a tarefa 4, uma vez que 20% dos participantes classificou a tarefa como sendo “Difícil”. Pelo referido, é possível concluir que o botão utilizado para ativar o mecanismo de resumo de *threads* através de palavras-chave necessita de ser melhorado. É ainda possível concluir que todos os outros componentes de interface são intuitivos e fáceis de utilizar, uma vez que todas as outras tarefas apresentaram níveis de dificuldade de conclusão bastante baixos.

Continuando a análise é possível concluir também que as funcionalidades de visualização orientada à conversa, extração de palavras-chave para resumo de *threads*, classificação de *threads* em trabalho ou lazer e de filtragem de mensagens são de utilidade bastante elevada (praticamente inquestionável), uma vez que 90%, 100%, 90% e 95% dos participantes considerou que as referidas funcionalidades eram pelo menos úteis, respetivamente. Pode também concluir-se que as funcionalidades referidas anteriormente estão bem conseguidas no Spidious, uma vez que 100%, 90%, 90%, 95% dos participantes, respetivamente, consideram que o resultado final das mesmas é pelo menos “Bom”. Aliás segundo os resultados obtidos na [Subsecção 6.2.2](#) pode concluir-se que o classificador automático do Spidious é de elevada qualidade, uma vez que apresentou uma

percentagem de acerto de 80% nas sessões de teste efetuadas. É ainda importante referir que 95% dos inquiridos considera que a interface do Spidious, de um modo geral, é pelo menos “Fácil” de ser utilizada. Considerando ainda algumas das opiniões relatadas na [Tabela 6.1](#), pode concluir-se que a interface do Spidious é objetiva, intuitiva e de fácil utilização.

No entanto, recorrendo aos resultados apresentados anteriormente, é possível também retirar mais algumas conclusões. Segundo 100% dos participantes os novos métodos de visualização, os algoritmos de aprendizagem, e o facto de estas duas componentes (visualização e algoritmos) estarem conjugados num cliente de e-mail web, o Spidious, aumentam claramente a eficiência de utilização face aos clientes de e-mail que os participantes utilizam habitualmente. Segundo os restantes resultados, é possível ainda concluir que a maior parte dos participantes estaria disposto a utilizar o Spidious no futuro (95%), mas não a pagar por ele (85%) (pelo menos em regime de mensalidade). No entanto os 15% de participantes que afirmaram poder pagar uma mensalidade pelo produto não podem ser ignorados, isto porque este é um mercado gigantesco. Aliás segundo a Radicatti Group [[Sar14](#)] este é um mercado que poderá chegar aos 5.2 biliões até 2018. Através destes últimos dados é possível concluir que o Spidious revelou-se um produto de verdadeiro sucesso, que teria grande probabilidade de singrar no mercado.

6.4 Resumo

Por fim, para terminar este capítulo, apresentar-se-á um resumo sob a forma de listagem, de toda a informação importante referida anteriormente ao longo deste capítulo. Com o referido resumo pretende-se fornecer ao leitor, uma forma rápida de entender os pontos fulcrais deste estudo realizado com os utilizadores.

- O botão que ativa o mecanismo de resumo de *threads* através de palavras-chave necessita de ser melhorado, uma vez que 20% dos inquiridos sentiu que executar a tarefa 4, que consistia em ativar as palavras-chave de resumo, foi “Difícil”.
- As funcionalidades de visualização orientada à conversa, extração de palavras-chave para resumo de *threads*, classificação de *threads* em trabalho ou lazer e de filtragem de mensagens são de utilidade bastante elevada, praticamente inquestionável.
- As funcionalidades de visualização orientada à conversa, extração de palavras-chave para resumo de *threads*, classificação de *threads* em trabalho ou lazer e de filtragem de mensagens demonstram um nível de execução bastante elevado no Spidious, praticamente inquestionável.
- O classificador automático do Spidious é de elevada qualidade, uma vez que apresentou uma percentagem de acerto de 80% nas sessões de teste efetuadas.
- A interface do Spidious é toda ela, de um modo geral, objetiva, intuitiva e de fácil utilização.

Spidious: Estudo com os utilizadores

- Os novos métodos de visualização, os algoritmos de aprendizagem, e o facto de estas duas componentes (visualização e algoritmos) estarem conjugados num cliente de e-mail web, o Spidious, aumentam claramente a eficiência de utilização face aos clientes de e-mail que os participantes utilizam habitualmente.
- A maior parte dos participantes estaria disposto a utilizar o Spidious no futuro (95%), mas não a pagar por ele (85%) (pelo menos em regime de mensalidade). No entanto os 15% de pessoas disponíveis para pagar pelo produto não pode ser ignorado, uma vez que este é um mercado gigantesco.
- O Spidious revelou-se um produto de verdadeiro sucesso, que teria grande probabilidade de singrar no mercado.

Spidious: Estudo com os utilizadores

Capítulo 7

Conclusões

Por fim, e a terminar esta dissertação, seguem-se as conclusões de todo o trabalho efetuado. As conclusões retiradas são de todo importantes porque representam parte do conhecimento gerado, o qual é esperado que permaneça válido por bastante tempo.

Ao longo do trabalho apresentado foram definidas as contribuições, os objetivos, e algumas questões de pesquisa (Capítulo 1). Foi também efetuada uma pequena apresentação sobre alguns fundamentos importantes e necessários para a compreensão deste trabalho sobre o e-mail (Capítulo 2). Apresentou-se também um estudo do estado da arte (Capítulo 3), um inquérito à Universidade do Porto (Capítulo 4), uma solução para o problema (Capítulo 5) e um estudo com utilizadores para validação da solução (Capítulo 6).

O presente capítulo apresentará de seguida um resumo das conclusões que foram possíveis obter (Secção 7.1), uma pequena abordagem às contribuições obtidas (Secção 7.2) e ainda por fim, uma breve referência ao trabalho a efetuar no futuro (Secção 7.3).

7.1 Resumo

O trabalho apresentado foi, dentro do possível, longo e aprofundado. Pese embora o facto de este ser um tema à volta do qual existe muita informação, foram selecionadas e analisadas diversas publicações, sendo algumas delas já bastante antigas. O mais notável foi perceber que a maior parte dos problemas a resolver no e-mail estão estudados e referidos há muito tempo (mais concretamente há cerca de 20 anos, como foi referido na Secção 1.2), mas que também grande parte dos mesmos não se encontram ainda resolvidos. O problema a tratar, referido na Secção 5.1, é extenso e de difícil resolução. Contudo esta dissertação pretendeu abordá-lo e solucioná-lo, ou pelo menos contribuir positivamente para a sua resolução.

A abordagem seguida por esta dissertação, para a resolução do problema, baseou-se numa solução que pretendeu principalmente atacar o problema com uma nova visualização, complementada por uma classificação automática, efetuada por algoritmos de aprendizagem supervisionada baseada em ações do utilizador, sendo estas duas componentes apresentadas sob a forma de um cliente de e-mail web.

Conclusões

O leque de conclusões que se poderia retirar é vasto, mas em primeira instância, proceder-se-á à análise das questões de pesquisa apresentadas, uma vez que foram estas às quais esta dissertação se propôs a responder. As questões de pesquisa apresentadas na [Secção 1.3](#) foram as seguintes:

1. Será que é possível melhorar a eficiência do utilizador no e-mail usando novos métodos e abordagens visuais?
2. Será que é possível melhorar a eficiência do utilizador no e-mail aplicando algoritmos de aprendizagem às contas de e-mail dos utilizadores?
3. Será que é possível melhorar a eficiência do utilizador no e-mail conjugando as abordagens anteriormente referidas num cliente de e-mail web?

Com recurso aos resultados apresentados e discutidos no [Capítulo 6](#), que divulga as informações obtidas do estudo com utilizadores realizado, foi possível responder às referidas questões de pesquisa. A resposta à primeira é totalmente afirmativa: sim, é possível melhorar a eficiência do utilizador no e-mail utilizando os novos métodos e abordagens visuais apresentados pelo Spidious. No estudo efetuado, 100% dos participantes responderam afirmativamente a esta pergunta depois de terem experimentado o Spidious.

Para a segunda questão de pesquisa a resposta é também ela afirmativa: sim, é possível melhorar a eficiência do utilizador no e-mail aplicando algoritmos de aprendizagem às contas de e-mail dos utilizadores. Mais uma vez a resposta a esta pergunta baseia-se no estudo efetuado, no qual 100% dos participantes responderam também afirmativamente à pergunta depois de utilizarem o Spidious.

Para concluir esta discussão, a resposta à terceira questão de pesquisa é mais uma vez afirmativa: sim, é possível melhorar a eficiência do utilizador no e-mail, conjugando uma nova visualização e algoritmos de aprendizagem, num novo cliente de e-mail web, o Spidious. De novo, a resposta obtida baseia-se no estudo efetuado, no qual 100% dos participantes responderam afirmativamente a esta pergunta, depois de terem utilizado o Spidious.

Por fim, é possível apresentar algumas conclusões obtidas por esta dissertação, recorrendo às ilações retiradas do referido estudo com utilizadores ([Capítulo 6](#)). Pode ainda então concluir-se que:

- As funcionalidades desenvolvidas no Spidious são extremamente úteis e funcionais, e ajudam o utilizador no desempenho das suas tarefas diárias no e-mail.
- A interface do Spidious é limpa, intuitiva e funcional.
- O classificador automático desenvolvido é de elevada qualidade.
- O Spidious tem potencialidade para se tornar um produto de verdadeiro sucesso no mercado.

7.2 Contribuições

No início desta dissertação foram apresentados os objetivos e as contribuições (Secção 1.4). Na presente secção serão apresentadas, de forma resumida, as contribuições obtidas.

Por um lado, o principal objetivo deste trabalho foi o de criar uma abordagem que permitisse resolver, ou pelo menos melhorar, os problemas verificados no e-mail, sendo que essa mesma abordagem teria de ser adaptada às exigências e padrões atuais verificados. Por outro lado esperava-se que a principal contribuição deste trabalho fosse a construção de um cliente de e-mail web que fosse capaz de aumentar a eficiência do utilizador ao lidar com o e-mail.

Com recurso a todos os resultados (Capítulo 6) e conclusões, já apresentadas anteriormente (Secção 7.1), é possível constatar que os objetivos foram não só cumpridos, mas até mesmo ultrapassados e as contribuições obtidas condizem com as contribuições esperadas.

A concluir esta dissertação abordar-se-á, de seguida, o trabalho futuro que poderá ser realizado, para que seja possível dar continuidade a tudo que foi apresentada até agora.

7.3 Trabalho Futuro

O início deste capítulo resumiu, de forma breve, o longo trabalho efetuado nesta dissertação. No entanto, e como sempre na engenharia de software, nada está definitivamente acabado, bem pelo contrário: está sempre disponível para ser melhorado. Assim sendo foram identificadas algumas funcionalidades a investigar e a trabalhar no futuro, muito devido ao facto de o estudo com utilizadores (Capítulo 6) ter sido bastante rico em conclusões.

O trabalho futuro tem por base os aspetos negativos do Spidious, apontados pelos utilizadores na Tabela 6.1. Segundo esses aspetos menos positivos seria então necessário adicionar teclas de atalho ao cliente de e-mail, adicionar a funcionalidade de gestão de tarefas, adicionar uma barra de pesquisa, destacar ainda mais as palavras-chave de resumo, fazer com que por defeito nenhuma *thread* apareça expandida, melhorar o aspeto do botão que permite ligar e desligar as palavras-chave de resumo e ainda avaliar a utilidade de poder criar mais filtros personalizados.

O presente capítulo encerra esta dissertação. Porém, como disse um dia Romain Rolland “*Nada está verdadeiramente feito enquanto restar alguma coisa para fazer*”. E na minha sincera opinião haverá sempre alguma coisa mais para fazer.

Conclusões

Referências

- [APS10] Douglas Aberdeen, O Pacovsky e Andrew Slater. The learning behind gmail priority inbox. ... *2010 Workshop on Learning on ...*, pages 3–6, 2010. URL: http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/us/pubs/archive/36955.pdf.
- [AZK09] T Ayodele, Shikun Zhou e R Khusainov. Email Grouping and Summarization: An Unsupervised Learning Technique. In *Computer Science and Information Engineering, 2009 WRI World Congress on*, volume 5, pages 575–579, 2009. doi:10.1109/CSIE.2009.298.
- [BDHS03] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard e Ian Smith. Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03*, pages 345–352, New York, NY, USA, 2003. ACM. URL: <http://doi.acm.org/10.1145/642611.642672>, doi:10.1145/642611.642672.
- [BF92] N Borenstein e N Freed. MIME (Multipurpose Internet Mail Extensions): Mechanisms for Specifying and Describing the Format of Internet Message Bodies, 1992.
- [BF93] N Borenstein e N Freed. MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies, 1993.
- [BL12] A Borg e N Lavesson. E-mail Classification Using Social Network Information. In *Availability, Reliability and Security (ARES), 2012 Seventh International Conference on*, pages 168–173, 2012. doi:10.1109/ARES.2012.84.
- [BS02] Olle Bälter e Candace L Sidner. Bifrost Inbox Organizer: Giving Users Control over the Inbox. In *Proceedings of the Second Nordic Conference on Human-computer Interaction, NordiCHI '02*, pages 111–118, New York, NY, USA, 2002. ACM. URL: <http://doi.acm.org/10.1145/572020.572034>, doi:10.1145/572020.572034.
- [BT06] Victoria Bellotti e Jim Thornton. Managing Activities with TV-Acta: TaskVista and Activity-Centered Task Assistant. In *In Proceedings of the Second SIGIR Workshop on Personal Information Management (PIM, 2006)*.
- [BTC⁺07] Victoria Bellotti, Jim Thornton, Alvin Chin, Diane Schiano e Nathan Good. TV-ACTA: embedding an activity-centered interface for task management in email. *work*, 5:11, 2007.

REFERÊNCIAS

- [CDGV01] J J Cadiz, Laura Dabbish, Anoop Gupta e Gina D Venolia. Supporting Email Workflow. Technical Report MSR-TR-2001-88, Microsoft Research, 2001. URL: <http://research.microsoft.com/apps/pubs/default.aspx?id=69881>.
- [CKM02] Elisabeth Crawford, Judy Kay e Eric McCreath. An Intelligent Interface for Sorting Electronic Mail. In *Proceedings of the 7th International Conference on Intelligent User Interfaces, IUI '02*, pages 182–183, New York, NY, USA, 2002. ACM. URL: <http://doi.acm.org/10.1145/502716.502747>, doi:10.1145/502716.502747.
- [CNZ07] Giuseppe Carenini, Raymond T Ng e Xiaodong Zhou. Summarizing Email Conversations with Clue Words. In *Proceedings of the 16th International Conference on World Wide Web, WWW '07*, pages 91–100, New York, NY, USA, 2007. ACM. URL: <http://doi.acm.org/10.1145/1242572.1242586>, doi:10.1145/1242572.1242586.
- [CNZ08] Giuseppe Carenini, Raymond T Ng e Xiaodong Zhou. Summarizing Emails with Conversational Cohesion and Subjectivity. In Kathleen McKeown, Johanna D Moore, Simone Teufel, James Allan e Sadaoki Furui, editors, *ACL*, pages 353–361. The Association for Computer Linguistics, 2008. URL: <http://dblp.uni-trier.de/db/conf/acl/acl2008.html#CareniniNZ08>.
- [Coh95] William W Cohen. Fast effective rule induction. In *Proceedings of the twelfth international conference on machine learning*, pages 115–123, 1995.
- [Coh96] William W Cohen. Learning Rules that Classify E-Mail. In *Proceedings of the 1996 AAAI Spring Symposium on Machine Learning and Information Access*, 1996.
- [CORGC04] Simon H Corston-Oliver, Eric Ringger, Michael Gamon e Richard Campbell. Task-focused summarization of email. Association for Computational Linguistics, 2004. URL: <http://research.microsoft.com/apps/pubs/default.aspx?id=68963>.
- [Cro82] D Crocker. Standard for the Format of ARPA Internet Text Messages, 1982.
- [Cro09] D Crocker. Internet Mail Architecture. RFC 5598 (Informational), 2009. URL: <http://www.ietf.org/rfc/rfc5598.txt>.
- [DB01] Nicolas Ducheneaut e Victoria Bellotti. E-mail As Habitat: An Exploration of Embedded Personal Information Management. *interactions*, 8(5):30–38, 2001. URL: <http://doi.acm.org/10.1145/382899.383305>, doi:10.1145/382899.383305.
- [Den82] Peter J Denning. ACM President's Letter: Electronic Junk. *Commun. ACM*, 25(3):163–165, 1982. URL: <http://doi.acm.org/10.1145/358453.358454>, doi:10.1145/358453.358454.
- [Dix09] Alan Dix. Human-Computer Interaction. In LING LIU e M.TAMER ÖZSU, editors, *Encyclopedia of Database Systems SE - 192*, pages 1327–1331. Springer US, 2009. URL: http://dx.doi.org/10.1007/978-0-387-39940-9_192, doi:10.1007/978-0-387-39940-9_192.

REFERÊNCIAS

- [DKV99] Judith Donath, Karrie Karahalios e Fernanda B Viégas. Visualizing Conversation. In *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 2 - Volume 2*, HICSS '99, pages 2023—, Washington, DC, USA, 1999. IEEE Computer Society. URL: <http://dl.acm.org/citation.cfm?id=874069.875994>.
- [DLH02] Douglas K Van Duyne, James Landay e Jason I Hong. *The Design of Sites: Patterns, Principles, and Processes for Crafting a Customer-Centered Web Experience*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2002.
- [DLK06] Mark Dredze, Tessa Lau e Nicholas Kushmerick. Automatically Classifying Emails into Activities. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, IUI '06, pages 70–77, New York, NY, USA, 2006. ACM. URL: <http://doi.acm.org/10.1145/1111449.1111471>, doi:10.1145/1111449.1111471.
- [DXW04] Angelo Dalli, Yunqing Xia e Yorick Wilks. FASIL Email Summarisation System. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. URL: <http://dx.doi.org/10.3115/1220355.1220498>, doi:10.3115/1220355.1220498.
- [FB96a] N Freed e N Borenstein. Multipurpose Internet Mail Extensions (MIME) Part Five: Conformance Criteria and Examples, 1996.
- [FB96b] N Freed e N Borenstein. Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies, 1996.
- [FB96c] N Freed e N Borenstein. Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types, 1996.
- [FBGS06] Danyel Fisher, A J Brush, Eric Gleave e Marc A Smith. Revisiting Whittaker & Sidner's "Email Overload" Ten Years Later. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, CSCW '06, pages 309–312, New York, NY, USA, 2006. ACM. URL: <http://doi.acm.org/10.1145/1180875.1180922>, doi:10.1145/1180875.1180922.
- [FKP96] N Freed, J Klensin e J Postel. Multipurpose Internet Mail Extensions (MIME) Part Four: Registration Procedures, 1996.
- [GSG10] Ashish Gupta, Ramesh Sharda e Robert a. Greve. You've got email! Does it really matter to process emails now or later? *Information Systems Frontiers*, 13(5):637–653, June 2010. URL: <http://link.springer.com/10.1007/s10796-010-9242-4>, doi:10.1007/s10796-010-9242-4.
- [Gwi02a] Jacek Gwizdka. Reinventing the Inbox: Supporting the Management of Pending Tasks in Email. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, pages 550–551, New York, NY, USA, 2002. ACM. URL: <http://doi.acm.org/10.1145/506443.506476>, doi:10.1145/506443.506476.

REFERÊNCIAS

- [Gwi02b] Jacek Gwizdka. TaskView: Design and Evaluation of a Task-based Email Interface. In *Proceedings of the 2002 Conference of the Centre for Advanced Studies on Collaborative Research, CASCON '02*, pages 4—. IBM Press, 2002. URL: <http://dl.acm.org/citation.cfm?id=782115.782119>.
- [Gwi04] Jacek Gwizdka. Email Task Management Styles: The Cleaners and the Keepers. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems, CHI EA '04*, pages 1235–1238, New York, NY, USA, 2004. ACM. URL: <http://doi.acm.org/10.1145/985921.986032>, doi:10.1145/985921.986032.
- [JDW01] Thomas Jackson, Ray Dawson e Darren Wilson. The cost of email interruption, 2001. doi:10.1108/13287260180000760.
- [JDW03a] Thomas Jackson, Ray Dawson e Darren Wilson. Reducing the effect of email interruptions on employees. *International Journal of Information Management*, 23(1):55–65, 2003. URL: <http://www.sciencedirect.com/science/article/pii/S0268401202000683>, doi:[http://dx.doi.org/10.1016/S0268-4012\(02\)00068-3](http://dx.doi.org/10.1016/S0268-4012(02)00068-3).
- [JDW03b] Thomas W Jackson, Ray Dawson e Darren Wilson. Understanding Email Interaction Increases Organizational Productivity. *Commun. ACM*, 46(8):80–84, 2003. URL: <http://doi.acm.org/10.1145/859670.859673>, doi:10.1145/859670.859673.
- [Ker03] Bernard Kerr. Thread Arcs: An email thread visualization. In *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, pages 211–218, 2003. doi:10.1109/INFVIS.2003.1249028.
- [KV09] Thomas Karagiannis e Milan Vojnovic. Behavioral Profiles for Advanced Email Features. In *Proceedings of the 18th International Conference on World Wide Web, WWW '09*, pages 711–720, New York, NY, USA, 2009. ACM. URL: <http://doi.acm.org/10.1145/1526709.1526805>, doi:10.1145/1526709.1526805.
- [LZP⁺09] Shixia Liu, Michelle X Zhou, Shimei Pan, Weihong Qian, Weijia Cai e Xiaoxiao Lian. Interactive, Topic-based Visual Text Summarization and Analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09*, pages 543–552, New York, NY, USA, 2009. ACM. URL: <http://doi.acm.org/10.1145/1645953.1646023>, doi:10.1145/1645953.1646023.
- [Mac88] Wendy E Mackay. More Than Just a Communication System: Diversity in the Use of Electronic Mail. In *Proceedings of the 1988 ACM Conference on Computer-supported Cooperative Work, CSCW '88*, pages 344–353, New York, NY, USA, 1988. ACM. URL: <http://doi.acm.org/10.1145/62266.62293>, doi:10.1145/62266.62293.
- [MGT⁺87] Thomas W Malone, Kenneth R Grant, Franklyn A Turbak, Stephen A Brobst e Michael D Cohen. Intelligent Information-sharing Systems. *Commun. ACM*, 30(5):390–402, 1987. URL: <http://doi.acm.org/10.1145/22899.22903>, doi:10.1145/22899.22903.
- [MK04] Mirko Mandic e Andruid Kerne. Familiar & intimacy-based email visualization. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages p14–p14. IEEE, 2004.

REFERÊNCIAS

- [MMF⁺02] Paul Moody, Paul Moody, Danyel Fisher, Danyel Fisher e Danyel Fisher. Studies of Automated Collection of Email Records. Technical report, University of Irvine, 2002. URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.59.504>.
- [Moo93] K Moore. MIME (Multipurpose Internet Mail Extensions) Part Two: Message Header Extensions for Non-ASCII Text, 1993.
- [Moo96] K Moore. MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text, 1996.
- [Moo03] Paul B Moody. Reinventing Email. Technical report, Citeseer, 2003.
- [MTK01] Smaranda Muresan, Evelyne Tzoukermann e Judith L Klavans. Combining Linguistic and Machine Learning Techniques for Email Summarization. In *Proceedings of the 2001 Workshop on Computational Natural Language Learning - Volume 7*, ConLL '01, pages 19:1—19:8, Stroudsburg, PA, USA, 2001. Association for Computational Linguistics. URL: <http://dx.doi.org/10.3115/1117822.1117837>, doi:10.3115/1117822.1117837.
- [NB03] Paula S Newman e John C Blitzer. Summarizing Archived Discussions: A Beginning. In *Proceedings of the 8th International Conference on Intelligent User Interfaces*, IUI '03, pages 273–276, New York, NY, USA, 2003. ACM. URL: <http://doi.acm.org/10.1145/604045.604097>, doi:10.1145/604045.604097.
- [New02] Paula S Newman. Email Archive Overviews Using Subject Indexes. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, pages 652–653, New York, NY, USA, 2002. ACM. URL: <http://doi.acm.org/10.1145/506443.506529>, doi:10.1145/506443.506529.
- [PD11] Larry L Peterson e Bruce S Davie. *Computer Networks, Fifth Edition: A Systems Approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 5th edition, 2011.
- [RGM⁺04] Steven L Rohall, Dan Gruen, Paul Moody, Martin Wattenberg, Mia Stern, Bernard Kerr, Bob Stachel, Kushal Dave, Robert Armes e Eric Wilcox. ReMail: A Reinvented Email Prototype. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 791–792, New York, NY, USA, 2004. ACM. URL: <http://doi.acm.org/10.1145/985921.985938>, doi:10.1145/985921.985938.
- [RGMK01] Steven L Rohall, Daniel Gruen, Paul Moody e Seymour Kellerman. Email visualizations to aid communications. In *Proceedings of InfoVis 2001 The IEEE Symposium on Information Visualization, IEEE*, volume 12, page 15, 2001.
- [Roh02] Steven L Rohall. Redesigning Email for the 21 st Century Workshop Position Paper. In *CSCW 2002 Workshop: Redesigning Email for the 21st Century*. New Orleans, LA, 2002.
- [RSCL04] Owen Rambow, Lokesh Shrestha, John Chen e Chirsty Lauridsen. Summarizing Email Threads. In *Proceedings of HLT-NAACL 2004: Short Papers*, HLT-NAACL-Short '04, pages 105–108, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. URL: <http://dl.acm.org/citation.cfm?id=1613984.1614011>.

REFERÊNCIAS

- [Sal91] Gerard Salton. Developments in automatic text retrieval. *Science*, 253(5023):974–980, 1991.
- [Sar14] Sara Radicati. Email Statistics Report, 2014-2018. Technical report, The Radicati Group, Inc., 2014.
- [SH02] S Sudarsky e R Hjelsvold. Visualizing electronic mail. In *Information Visualization, 2002. Proceedings. Sixth International Conference on*, pages 3–9, 2002. doi:10.1109/IV.2002.1028749.
- [SK99] Richard B Segal e Jeffrey O Kephart. MailCat: An Intelligent Assistant for Organizing e-Mail. In *Proceedings of the Third Annual Conference on Autonomous Agents, AGENTS '99*, pages 276–282, New York, NY, USA, 1999. ACM. URL: <http://doi.acm.org/10.1145/301136.301209>, doi:10.1145/301136.301209.
- [SMR07] Lokesh Shrestha, Kathleen Mckeown e Owen Rambow. Using question-answer pairs in extractive summarization of email conversations. In *In Proceedings of CILing, volume 4394 of Lecture Notes in Computer Science*, pages 542–550, 2007.
- [TT03] JR Tyler e JC Tang. When can I expect an email response? A study of rhythms in email usage. *ECSCW 2003*, pages 14–18, 2003. URL: http://link.springer.com/chapter/10.1007/978-94-010-0068-0_13, doi:10.1007/978-94-010-0068-0_13.
- [UCMN09] Jan Ulrich, Giuseppe Carenini, Gabriel Murray e Raymond Ng. Regression-Based Summarization of Email Conversations, 2009. URL: <http://aaai.org/ocs/index.php/ICWSM/09/paper/view/188>.
- [Vac14] M Vacek. How to survive email. In *Applied Computational Intelligence and Informatics (SACI), 2014 IEEE 9th International Symposium on*, pages 49–54, 2014. doi:10.1109/SACI.2014.6840097.
- [VN03] Gina Danielle Venolia e Carman Neustaedter. Understanding Sequence and Reply Relationships Within Email Conversations: A Mixed-model Visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03*, pages 361–368, New York, NY, USA, 2003. ACM. URL: <http://doi.acm.org/10.1145/642611.642674>, doi:10.1145/642611.642674.
- [VPH10] Romain Vuillemot, Jean-Marc Petit e Mohand-Saïd Hacid. Shift-BOX: INBOX Time Shifting to Reduce Email Clutter. In *Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*, 2010.
- [WBM05] Steve Whittaker, Victoria Bellotti e Paul Moody. Introduction to This Special Issue on Revisiting and Reinventing e-Mail. *Hum.-Comput. Interact.*, 20(1):1–9, 2005. URL: http://dx.doi.org/10.1207/s15327051hci2001&2_1, doi:10.1207/s15327051hci2001&2_1.
- [WM04] Stephen Wan e Kathy McKeown. Generating Overview Summaries of Ongoing Email Thread Discussions. In *Proceedings of the 20th International Conference on Computational Linguistics, COLING '04*, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. URL: <http://dx.doi.org/10.3115/1220355.1220434>, doi:10.3115/1220355.1220434.

REFERÊNCIAS

- [WS96] Steve Whittaker e Candace Sidner. Email overload: exploring personal information management of email. In *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96*, pages 276–283, 1996. URL: <http://portal.acm.org/citation.cfm?doid=238386.238530>, doi:10.1145/238386.238530.
- [YBSL97] Kelvin S Yiu, Ronald Baecker, Nancy Silver e Byron Long. A Time-Based Interface for Electronic Mail and Task Management. *Design of Computing Systems: Proceedings of HCI International*, 2:19–22, 1997.
- [Yi07] Zhixian Yi. E-Mail Use Patterns in the United States. *Chinese Librarianship: An International Electronic Journal*, Dec2007(24):1–11p, 2007. URL: <http://www.white-clouds.com/iclc/cliej/cl24yi.pdf>.

REFERÊNCIAS

Anexo A

Inquérito sobre o e-mail aos utilizadores

Spidious - Pesquisa com os utilizadores

Prezado(a) colega, este inquérito tem como objectivo principal perceber de que forma os utilizadores interagem com o seu cliente de e-mail atual e se encontram, ou não, satisfeitos com o mesmo. Paralelamente pretende ainda perceber se os utilizadores se demonstram recetivos a novas formas de visualização e à aplicação de técnicas de aprendizagem ao e-mail.

Os resultados provenientes deste inquérito serão utilizados na minha dissertação de mestrado que tem o título: "Spidious: Improving User's Efficiency on Email with Machine Learning and a Better Visualization Approach".

Ao responder a este inquérito estará certamente a contribuir para a evolução de um dos meios de comunicação digital mais usado em todo o mundo.

Se tiver algum tipo de dúvida a responder a este inquérito, não hesite em me contactar:

Tiago Costa - ei10114@fe.up.pt

* Required

Geral

1. **Sexo:** *

Mark only one oval.

- Feminino
 Masculino

2. **Qual a sua ocupação profissional atualmente?** *

Mark only one oval.

- Estudante
 Trabalhador-Estudante
 Trabalhador
 Other:

E-mail

3. **Que tipo de cliente de e-mail utiliza predominantemente?** *

Mark only one oval.

- Desktop
 Mobile
 Web

Inquérito sobre o e-mail aos utilizadores

4. Da seguinte lista, quais dos clientes de e-mail utiliza? *

Check all that apply.

- OSX Mail
- iOS Mail
- Outlook Desktop
- Outlook Web
- Outlook Mobile
- Gmail Web
- Gmail Mobile
- Inbox Web
- Inbox Mobile
- Mailbox OS X
- Mailbox Mobile
- Thunderbird
- FEUP Webmail
- Other:

5. Onde costuma consultar o seu e-mail? *

Mark only one oval.

- Casa
- Faculdade / Escritório
- Ambos
- Other:

6. Qual a frequência com a qual verifica o seu e-mail? *

Mark only one oval.

- Menos que 1 vez / dia
- Entre 1 a 5 vezes / dia
- Menos que 1 vez / hora
- 1 ou mais vezes / hora

7. Qual o tempo médio que despende diariamente com o seu e-mail? *

Mark only one oval.

- Menos de 30 minutos
- Entre 30 minutos a 1 hora
- Entre 1 a 3 horas
- Mais de 3 horas

8. Qual o número de mensagens de e-mail que recebe por dia? *

Mark only one oval.

- Menos de 10
- Entre 10 e 20
- Entre 21 e 50
- Mais de 50

9. Qual o número de mensagens de e-mail que envia por dia? *

Mark only one oval.

- Menos de 10
- Entre 10 e 20
- Entre 21 e 50
- Mais de 50

Visualização

10. Qual o número médio de mensagens não lidas que a sua caixa de entrada apresenta normalmente? *

Mark only one oval.

- Menos de 10
- Entre 10 e 50
- Entre 51 e 100
- Mais de 100

11. Que tipo de mensagens deixa na sua caixa de entrada? *

Check all that apply.

- Mensagens com conversas em curso
- Mensagens não arquivadas
- Tarefas incompletas
- Other:

12. Utiliza normalmente pastas, que não as existentes por defeito, para organizar o seu email? Se sim, qual a sua quantidade(incluindo subpastas)?

Mark only one oval.

- Menos de 2
- Entre 2 e 5
- Entre 6 e 10
- Mais de 11

Inquérito sobre o e-mail aos utilizadores

13. **Qual o método de pesquisa que usa mais frequentemente quando procura uma mensagem no seu e-mail? ***

Mark only one oval.

- Barra de pesquisa
 Ordenação de mensagens
 Pesquisa visual
 Other:

14. **Ao efetuar uma pesquisa, qual o critério de pesquisa ao qual recorre mais vezes? ***

Mark only one oval.

- Remetente
 Data de envio
 Assunto
 Corpo da mensagem
 Other:

15. **Utiliza algum tipo de notificação automática aquando da chegada de um novo e-mail? Se sim, qual? ***

Mark only one oval.

- Auditiva
 Visual
 Ambas
 Nenhuma

16. **Considera as notificações automáticas de e-mail intrusivas? ***

Mark only one oval.

- Sim
 Não

17. **Classifique numa escala de 1 a 5, o quão satisfeito se encontra, numa perspectiva de utilizador, com a atual interface da generalidade dos clientes de e-mail. ***

Mark only one oval.

1 2 3 4 5

Muito insatisfeito Muito satisfeito

18. **Quando necessita de escrever um novo e-mail, é frequente servir-se de uma mensagem anterior do destinatário pretendido, para assim evitar escrevê-lo de novo? Classifique numa escala de 1 a 5. ***

Mark only one oval.

1 2 3 4 5

Pouco Frequente Muito Frequente

19. **Sente que, de alguma forma, o atual paradigma de visualização no e-mail afeta a sua eficiência e produtividade enquanto utilizador? ***

Mark only one oval.

- Sim
 Não

Inteligência

20. **Está satisfeito com as funcionalidades de inteligência artificial e aprendizagem já desenvolvidas para o e-mail (como por exemplo filtros)? ***

Mark only one oval.

- Sim
 Não

21. **Sente que, de alguma forma, recebe e-mails que não têm interesse em ser lidos no momento em que os recebe? ***

Mark only one oval.

- Sim
 Não

22. **Enquanto utilizador do e-mail, sente-se sobrecarregado com a quantidade de e-mails que tem de processar? ***

Mark only one oval.

- Sim
 Não

23. **Gostaria de utilizar um cliente de e-mail que fosse capaz de se adaptar ao seu perfil de utilização, e aprender quais os e-mails importantes para si a determinada hora do dia (ou seja, que tivesse em atenção o contexto de utilização)? ***

Mark only one oval.

- Sim
 Não

24. **Considera útil perceber o conteúdo de um dado conjunto de mensagens de e-mail, antes mesmo de o ler, através da visualização de meta-informação? ***

Mark only one oval.

- Sim
 Não

25. **Considera que uma classificação clara das mensagens de e-mail em trabalho ou lazer pode contribuir para um aumento da sua produtividade e eficiência ao utilizar o e-mail?**

*

Mark only one oval.

Sim

Não

Considerações Finais

26. **Que problemas encontra nos clientes de e-mail atuais?**

.....
.....
.....
.....
.....

27. **Que funcionalidades gostaria de ver implementadas num novo cliente de e-mail?**

.....
.....
.....
.....
.....

Inquérito sobre o e-mail aos utilizadores

Anexo B

Spidious: Estudo com os utilizadores

B.1 Pré-Questionário

Spidious - Pré-questionário do estudo com utilizadores

Prezado(a) colega, este estudo com utilizadores tem como objectivo principal aferir se as novas funcionalidades do Spidious apresentam de facto melhorias efetivas para os utilizadores em relação aos clientes de e-mail atuais. Este pré-questionário irá essencialmente servir para perceber os hábitos atuais do utilizador com o seu cliente de e-mail, e se, antes de mesmo de experimentar, antevê vantagens e potencialidades nas funcionalidades do Spidious.

Os resultados provenientes deste estudo com utilizadores serão utilizados na minha dissertação de mestrado que tem o título: "Spidious: Improving User's Efficiency on Email with Machine Learning and a Better Visualization Approach".

Ao participar neste estudo estará certamente a contribuir para a evolução de um dos meios de comunicação digital mais usado em todo o mundo.

* Required

1. Identificar Único *

.....

E-mail

2. Que tipo de cliente de e-mail utiliza predominantemente? *

Mark only one oval.

- Desktop
- Mobile
- Web

3. Da seguinte lista, quais dos clientes de e-mail utiliza? *

Check all that apply.

- OSX Mail
- iOS Mail
- Outlook Desktop
- Outlook Web
- Outlook Mobile
- Gmail Web
- Gmail Mobile
- Inbox Web
- Inbox Mobile
- Mailbox OS X
- Mailbox Mobile
- Thunderbird
- FEUP Webmail
- Other:

4. Onde costuma consultar o seu e-mail? *

Mark only one oval.

- Casa
- Faculdade / Escritório
- Ambos
- Other:

5. Qual a frequência com a qual verifica o seu e-mail? *

Mark only one oval.

- Menos que 1 vez / dia
- Entre 1 a 5 vezes / dia
- Menos que 1 vez / hora
- 1 ou mais vezes / hora

6. Qual o tempo médio que despende diariamente com o seu e-mail? *

Mark only one oval.

- Menos de 30 minutos
- Entre 30 minutos a 1 hora
- Entre 1 a 3 horas
- Mais de 3 horas

7. Qual o número de mensagens de e-mail que recebe por dia? *

Mark only one oval.

- Menos de 10
- Entre 10 e 20
- Entre 21 e 50
- Mais de 50

8. Qual o número de mensagens de e-mail que envia por dia? *

Mark only one oval.

- Menos de 10
- Entre 10 e 20
- Entre 21 e 50
- Mais de 50

Visualização

9. Qual o número médio de mensagens não lidas que a sua caixa de entrada apresenta normalmente? *

Mark only one oval.

- Menos de 10
- Entre 10 e 50
- Entre 51 e 100
- Mais de 100

10. Que tipo de mensagens deixa na sua caixa de entrada? *

Check all that apply.

- Mensagens com conversas em curso
- Mensagens não arquivadas
- Tarefas incompletas
- Other:

11. Utiliza normalmente pastas, que não as existentes por defeito, para organizar o seu email? Se sim, qual a sua quantidade(incluindo subpastas)?

Mark only one oval.

- Menos de 2
- Entre 2 e 5
- Entre 6 e 10
- Mais de 11

12. Qual o método de pesquisa que usa mais frequentemente quando procura uma mensagem no seu e-mail? *

Mark only one oval.

- Barra de pesquisa
- Ordenação de mensagens
- Pesquisa visual
- Other:

13. Ao efetuar uma pesquisa, qual o critério de pesquisa ao qual recorre mais vezes? *

Mark only one oval.

- Remetente
- Data de envio
- Assunto
- Corpo da mensagem
- Other:

14. Utiliza algum tipo de notificação automática aquando da chegada de um novo e-mail? Se sim, qual? *

Mark only one oval.

- Auditiva
- Visual
- Ambas
- Nenhuma

15. Considera as notificações automáticas de e-mail intrusivas? *

Mark only one oval.

- Sim
- Não

16. Classifique numa escala de 1 a 5, o quão satisfeito se encontra, numa perspectiva de utilizador, com a atual interface da generalidade dos clientes de e-mail. *

Mark only one oval.

	1	2	3	4	5	
Muito insatisfeito	<input type="radio"/>	Muito satisfeito				

17. Quando necessita de escrever um novo e-mail, é frequente servir-se de uma mensagem anterior do destinatário pretendido, para assim evitar escrevê-lo de novo? Classifique numa escala de 1 a 5. *

Mark only one oval.

	1	2	3	4	5	
Pouco Freqüente	<input type="radio"/>	Muito Freqüente				

18. Sente que, de alguma forma, o atual paradigma de visualização no e-mail afeta a sua eficiência e produtividade enquanto utilizador? *

Mark only one oval.

- Sim
 Não

19. Gostaria de ter um cliente de e-mail que mudasse o paradigma de visualização atual do e-mail, para um paradigma de conversação orientado ao remetente? *

Mark only one oval.

- Sim
 Não

20. Gostaria de ter um cliente de e-mail que permitisse que a resposta a uma mensagem de e-mail fosse tão simples como a resposta a um sistema de mensagens instantâneas? *

Mark only one oval.

- Sim
 Não

Inteligência

21. Está satisfeito com as funcionalidades de inteligência artificial e aprendizagem já desenvolvidas para o e-mail (como por exemplo filtros)? *

Mark only one oval.

- Sim
 Não

22. Sente que, de alguma forma, recebe e-mails que não têm interesse em ser lidos no momento em que os recebe? *

Mark only one oval.

- Sim
 Não

23. Enquanto utilizador do e-mail, sente-se sobrecarregado com a quantidade de e-mails que tem de processar? *

Mark only one oval.

- Sim
 Não

24. Gostaria de utilizar um cliente de e-mail que fosse capaz de se adaptar ao seu perfil de utilização, e aprender quais os e-mails importantes para si a determinada hora do dia (ou seja, que tivesse em atenção o contexto de utilização)? *

Mark only one oval.

- Sim
 Não

25. Considera útil perceber o conteúdo de um dado conjunto de mensagens de e-mail, antes mesmo de o ler, através da visualização de meta-informação? *

Mark only one oval.

- Sim
 Não

26. Considera que uma classificação clara das mensagens de e-mail em trabalho ou lazer pode contribuir para um aumento da sua produtividade e eficiência ao utilizar o e-mail? *

Mark only one oval.

- Sim
 Não

B.2 Resultados do Pré-Questionário

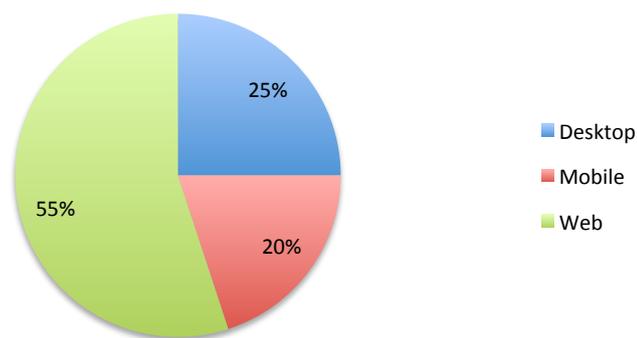


Figura B.1: Tipo de cliente de e-mail utilizado.

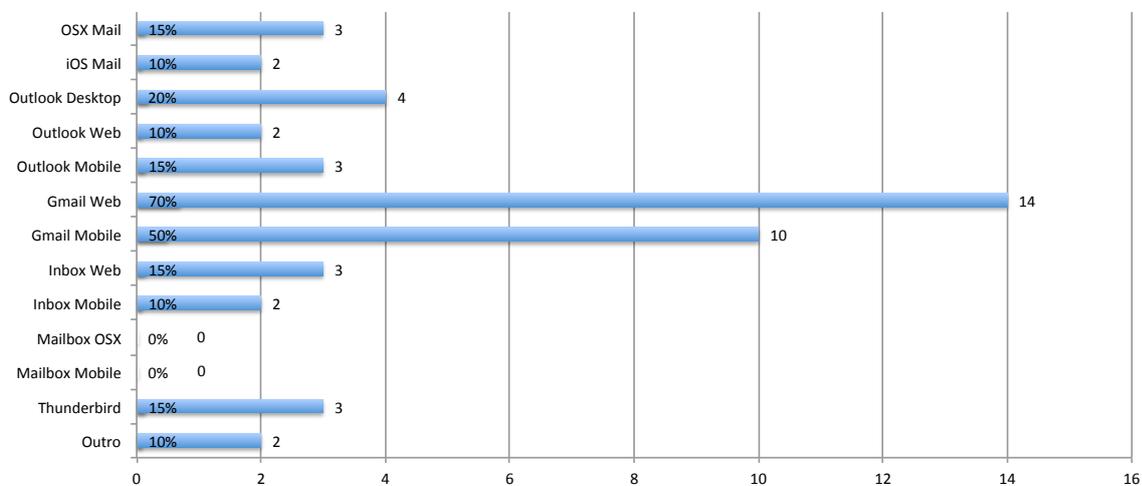


Figura B.2: Clientes de e-mail utilizados.

Spidious: Estudo com os utilizadores

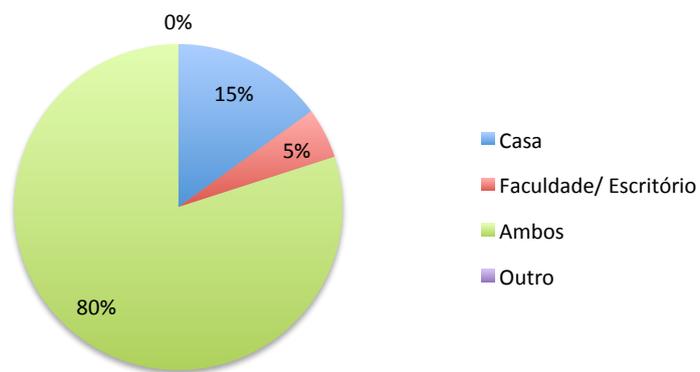


Figura B.3: Local de consulta do e-mail.

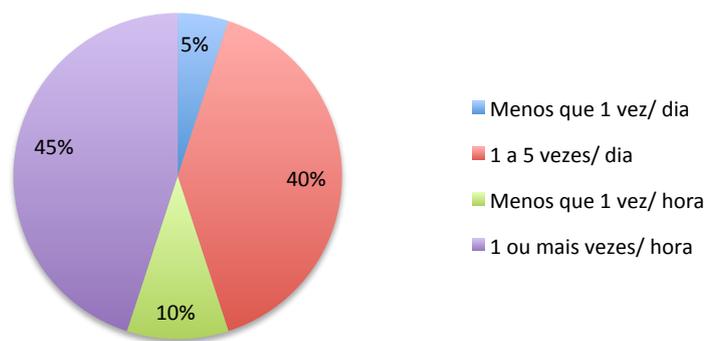


Figura B.4: Frequência de consulta do e-mail.

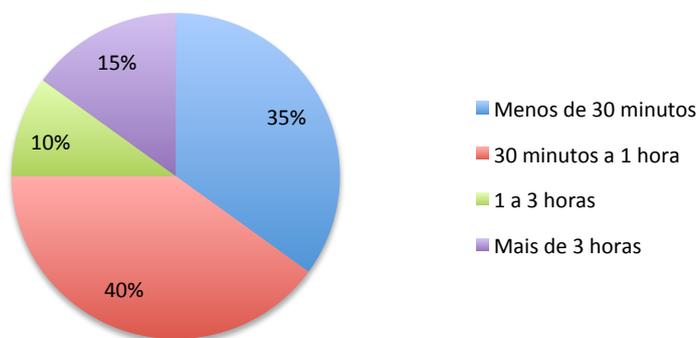


Figura B.5: Tempo médio diário despendido no e-mail.

Spidious: Estudo com os utilizadores

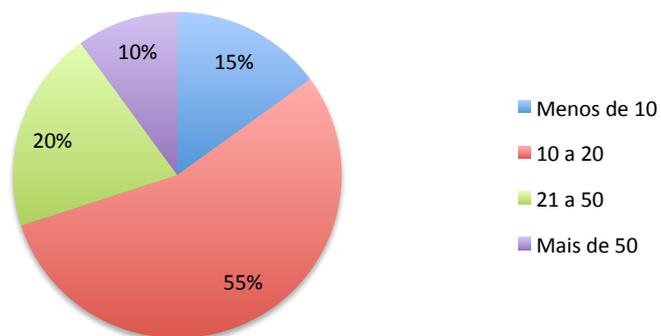


Figura B.6: Mensagens de e-mail recebidas por dia.

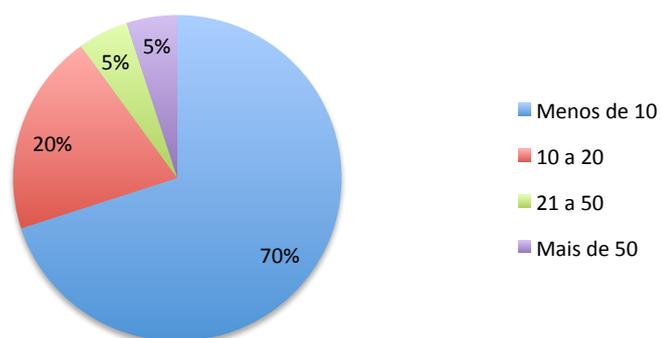


Figura B.7: Mensagens de e-mail enviadas por dia.

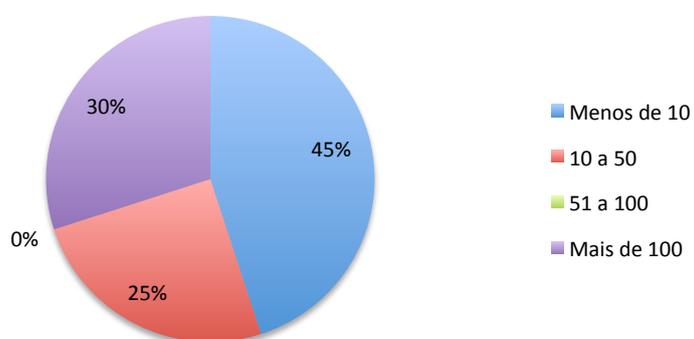


Figura B.8: Número médio de mensagens não lidas na caixa de entrada.

Spidious: Estudo com os utilizadores

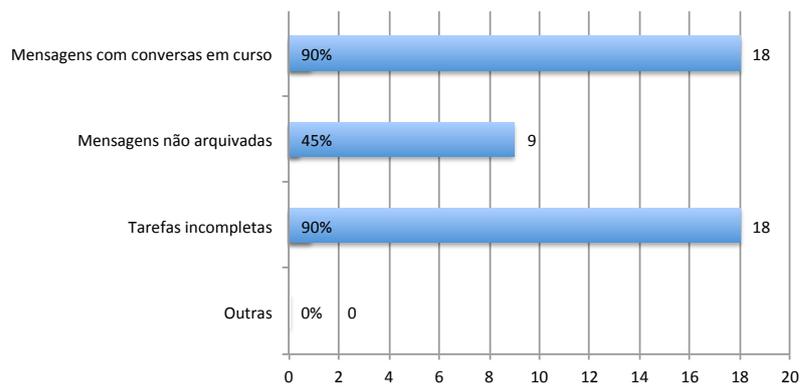


Figura B.9: Tipo de mensagens de e-mail deixadas na caixa de entrada.

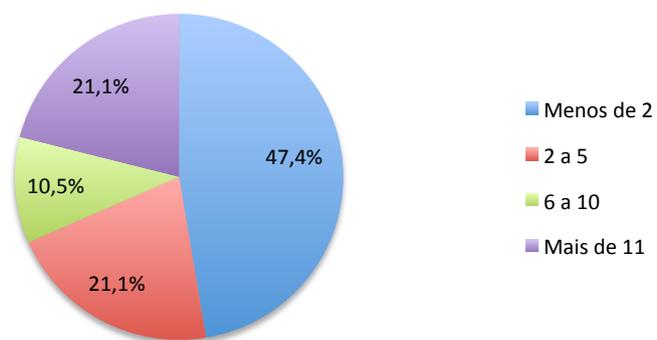


Figura B.10: Número de pastas excluindo as existentes por defeito.

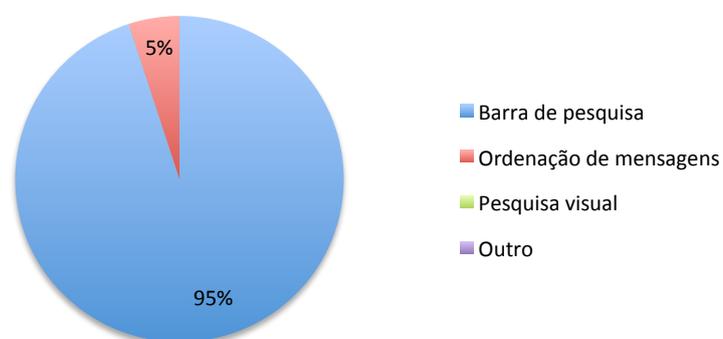


Figura B.11: Método de pesquisa de mensagens mais utilizado.

Spidious: Estudo com os utilizadores

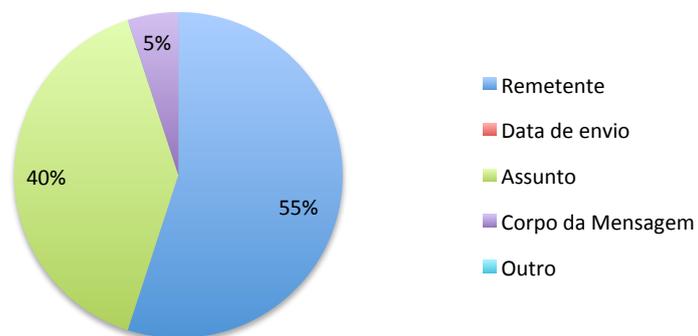


Figura B.12: Critério de pesquisa mais utilizado.

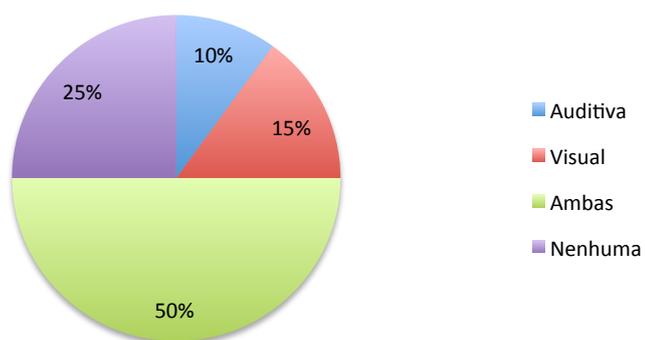


Figura B.13: Tipo de notificações automáticas utilizadas.

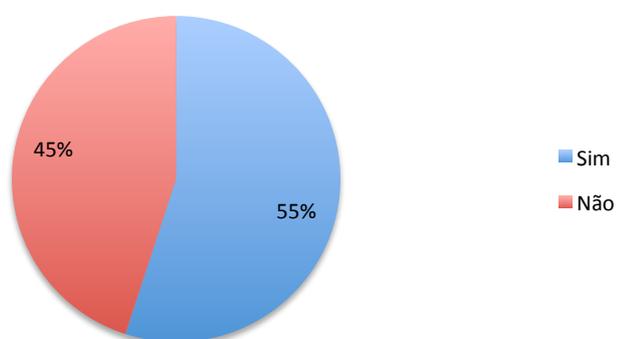


Figura B.14: Classificação das notificações automáticas quanto à sua intrusividade.

Spidious: Estudo com os utilizadores

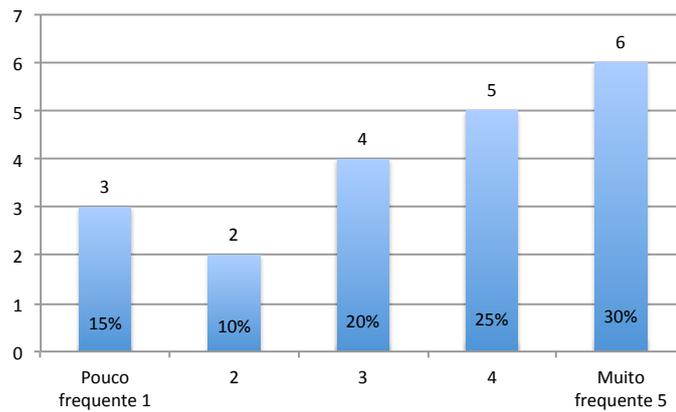


Figura B.15: Frequência de utilização de uma mensagem já recebida para obter o e-mail do destinatário para uma nova mensagem. $\bar{x} \approx 3.45$; $\sigma \approx 1.43$

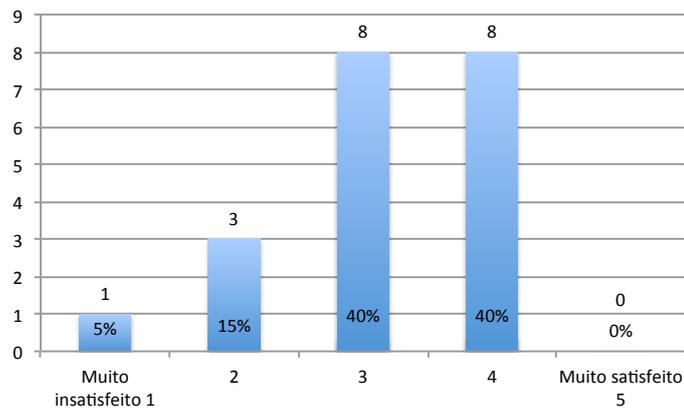


Figura B.16: Satisfação dos utilizadores em relação à atual interface dos clientes de e-mail. $\bar{x} \approx 3.15$; $\sigma \approx 0.88$

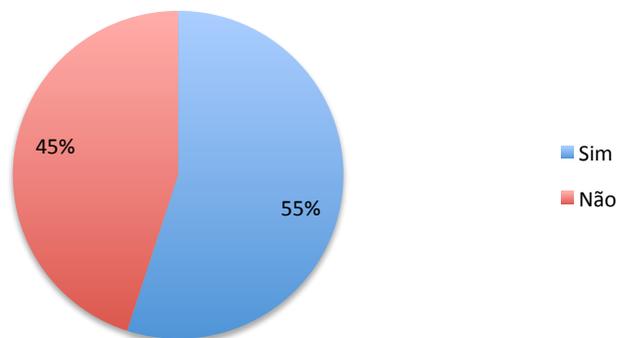


Figura B.17: Opinião dos utilizadores quanto à prejudicialidade da atual visualização do e-mail face à sua produtividade pessoal.

Spidious: Estudo com os utilizadores

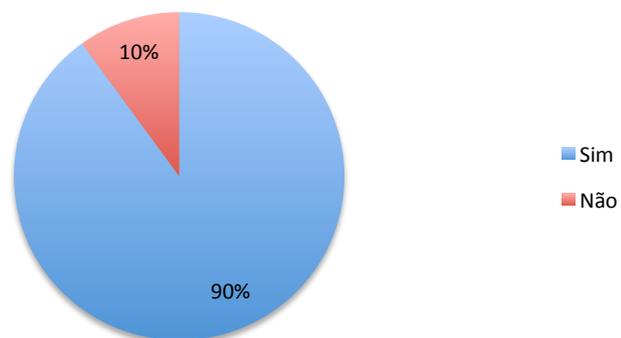


Figura B.18: Opinião dos utilizadores quanto à vontade de terem um cliente de e-mail com uma visualização orientada ao remetente.

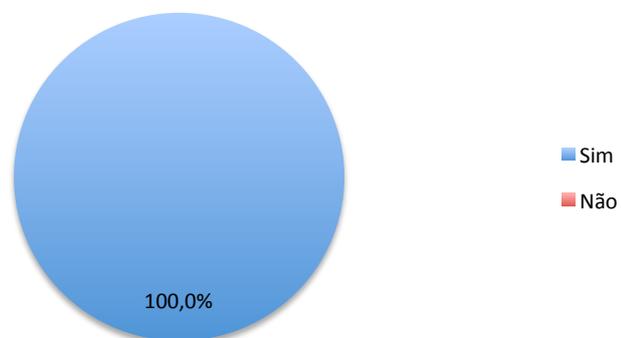


Figura B.19: Opinião dos utilizadores quanto à vontade de terem um cliente de e-mail que permita uma resposta rápida tal como acontece num sistema de mensagens instantâneas.

Spidious: Estudo com os utilizadores

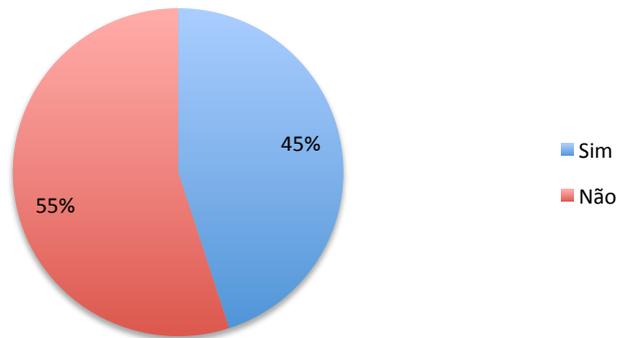


Figura B.20: Nível de satisfação com as funcionalidades de inteligência e aprendizagem já desenvolvidas para o e-mail.

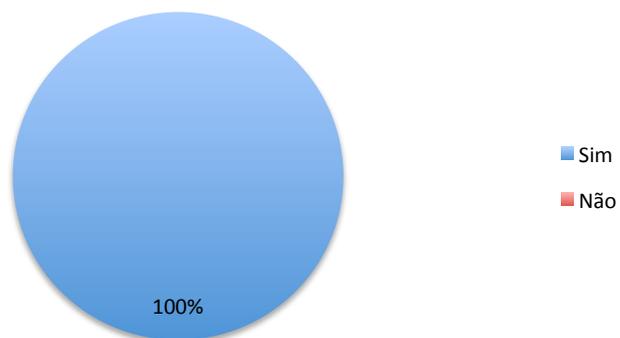


Figura B.21: Receção de e-mails sem interesse em ser lidos aquando da sua chegada.

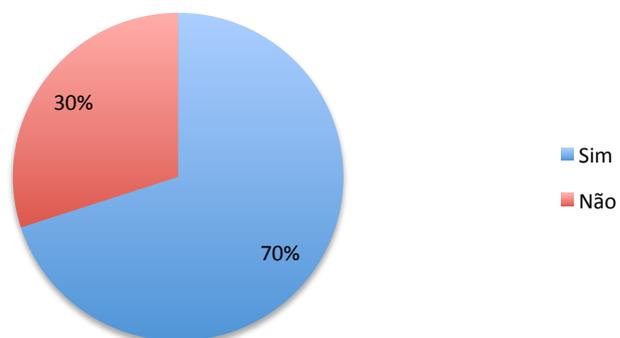


Figura B.22: Opinião sobre sobrecarga causada pela quantidade de mensagens de e-mail para processar.

Spidious: Estudo com os utilizadores

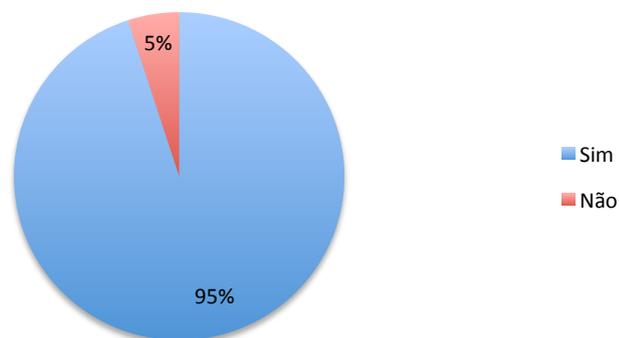


Figura B.23: Pretensão por um cliente de e-mail capaz de se adaptar ao perfil do utilizador.

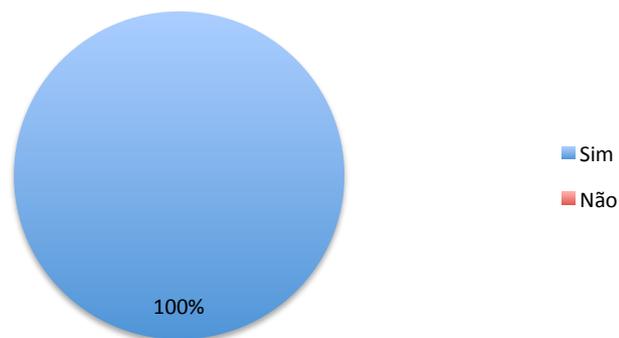


Figura B.24: Pretensão por perceber o conteúdo de um conjunto de mensagens antes mesmo de as ler através da visualização de meta-informação.

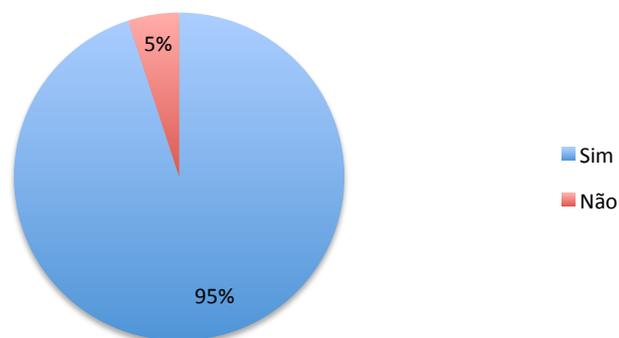


Figura B.25: Opinião sobre aumento de produtividade e eficiência no uso do e-mail causado por uma classificação clara das *threads* em trabalho ou lazer.

B.3 Guião de Tarefas

Spidious - Guião de Tarefas para o Estudo com Utilizadores

Prezado(a) colega, este guião apresenta um conjunto de tarefas e tem como objectivo principal permitir aos utilizadores experimentar todas as novas funcionalidades que o Spidious implementa para que, com isso, lhes seja possível perceber se o produto no seu todo acarreta, ou não, claras melhorias e vantagens.

Os resultados provenientes deste estudo com utilizadores serão utilizados na minha dissertação de mestrado que tem o título: "Spidious: Improving User's Efficiency on Email with Machine Learning and a Better Visualization Approach".

Ao participar neste estudo estará certamente a contribuir para a evolução de um dos meios de comunicação digital mais usado em todo o mundo.

* Required

Cenário

Considere que se chama Tiago Costa e é um estudante finalista no Mestrado em Engenharia Informática e Computação da Faculdade de Engenharia da Universidade do Porto. No âmbito da sua dissertação, o Tiago desenvolveu um cliente de e-mail web, de seu nome Spidious, que muda o paradigma comum aos clientes de e-mail atuais. De momento, o Tiago encontra-se a utilizar o Spidious no seu dia-a-dia e de entre as suas diversas conversações, uma delas é com Tim Bake o CEO da Pear Inc., a maior empresa de software do mundo, que está a negociar a compra do Spidious com o Tiago.

1. Identificador Único *

.....

Tarefa 1 - Classifique algumas Threads de E-mail

Efetue a classificação de 20 threads de e-mail diferentes, em Trabalho ou Lazer, tentando ser o mais diversificado possível ao longo da lista de conversações.

Após efetuar a classificação, o Spidious irá classificar as restantes mensagens com base no que aprendeu da sua classificação de exemplo.

2. Como classifica o sistema de classificação automática de threads do Spidious para os restantes e-mails? *

Mark only one oval.

	1	2	3	4	5	
Muito Mau	<input type="radio"/>	Muito Bom				

3. **Atribua uma classificação de Trabalho ou Lazer para à seguinte mensagem de e-mail: "Tiago your work is awesome! We'll buy your work by the value you want." ***

Mark only one oval.

- Trabalho
 Lazer

4. **Qual a classificação automática atribuída pelo Spidious à mensagem anterior? (A mensagem anterior faz parte de uma nova thread recebida na conversa com Tim Bake) ***

Mark only one oval.

- Trabalho
 Lazer

Tarefa 2 - Encontre a Thread

Encontre uma Thread que fala sobre o campeonato português de futebol, na conversa com o Tony Stacks.

5. **Quem iniciou esta Thread? ***

Mark only one oval.

- Tony Stacks
 Tiago Costa

6. **Quantas mensagens compõe esta Thread? ***

Mark only one oval.

- 0
 1
 2
 3

7. **Existem outras Threads na conversa com o Tony Stacks? ***

Mark only one oval.

- Sim
 Não

Tarefa 3 - Responda a uma mensagem

Na Thread que fala sobre o campeonato português de futebol, na conversa com o Tony Stacks, responda por favor à última mensagem deixada pelo Tony Stacks nessa Thread.

8. Com quantas mensagens ficou a referida Thread depois de efetuada a sua resposta? *

Mark only one oval.

- 1
 2
 3
 4

9. Gostou da interface de resposta a uma Thread que acabou de utilizar? *

Mark only one oval.

- Sim
 Não

Tarefa 4 - Obtenha as palavras-chave de resumo de Threads

Na conversação que tem sido usada para as tarefas anteriores, a conversação com o Tony Stacks, ative por favor agora o resumo de Threads.

Depois disso, confira se, para as duas Threads de conversa com o Tony Stacks, as palavras-chave geradas correspondem com o texto das mensagens trocadas nas respetivas Threads.

10. Qual o número médio de palavras-chave geradas para cada Thread? *

Mark only one oval.

- 0
 1
 2
 3
 4
 5

11. Acha que as palavras-chave geradas estão de acordo com o conteúdo presente em cada Thread? *

Mark only one oval.

- Sim
 Não

12. **Gostou da interface implementada para obter as palavras-chave de resumo das Threads? ***

Mark only one oval.

Sim

Não

13. **Achou a interface implementada para obter as palavras-chave de resumo das Threads intuitiva? ***

Mark only one oval.

Sim

Não

Tarefa 5 - Escreva uma Nova Thread

Na conversa que tem sido usada para as tarefas anteriores, a conversa com o Tony Stacks, crie agora uma nova Thread, com um assunto diferente dos anteriores.

14. **Com quantas Threads ficou a conversa com o Tony Stacks? ***

Mark only one oval.

1

2

3

15. **Gostou da interface implementada para criar uma nova Thread com um assunto diferente, dentro de uma conversa já criada? ***

Mark only one oval.

Sim

Não

Tarefa 6 - Botão de nova conversa

Clique agora no botão que, na sua opinião, serve para criar uma nova conversa.

16. **Achou o botão para criar uma nova conversa intuitivo? ***

Mark only one oval.

Sim

Não

Tarefa 7 - Filtre as mensagens de e-mail

Por fim, nesta última tarefa, aplique um filtro que seja capaz de apenas lhe mostrar as mensagens classificadas como Lazer.

17. Achou a interface de filtragem implementada intuitiva? *

Mark only one oval.

- Sim
- Não

Powered by
 Google Forms

B.4 Resultados do Guião de Tarefas

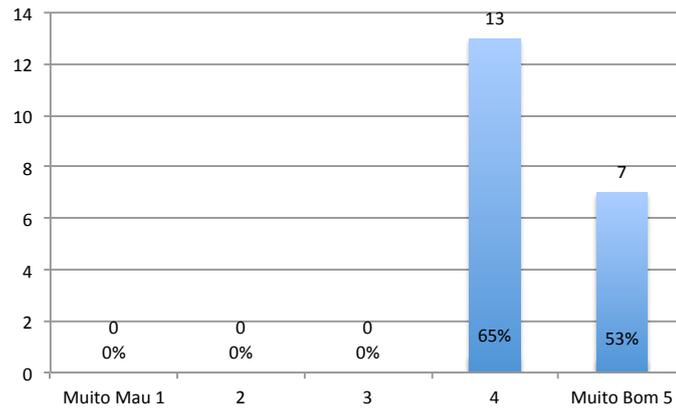


Figura B.26: Opinião sobre o sistema de classificação automática do Spidious para as restantes *threads* não classificadas pelo utilizador. $\bar{x} \approx 4.35$; $\sigma \approx 0.49$

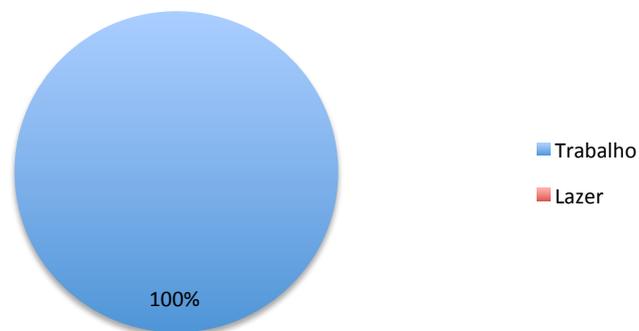


Figura B.27: Opinião sobre a classificação a atribuir à seguinte mensagem de e-mail: “*Tiago your work is awesome! We’ll buy your work by the value you want*”.

Spidious: Estudo com os utilizadores

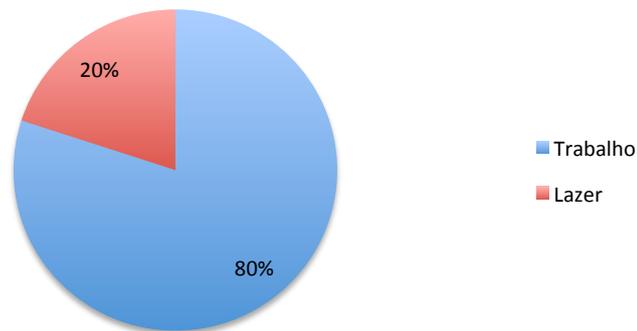


Figura B.28: Classificação automática atribuída pelo Spidious à mensagem anterior (A mensagem anterior faz parte de uma nova *thread* recebida na conversa com Tim Bake).

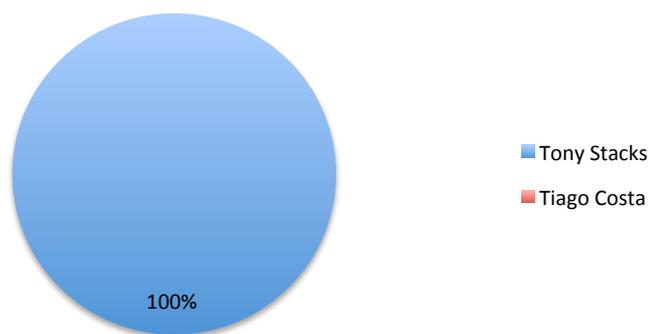


Figura B.29: Opinião sobre quem iniciou a *thread* que fala sobre o campeonato português de futebol na conversa com o Tony Stacks.

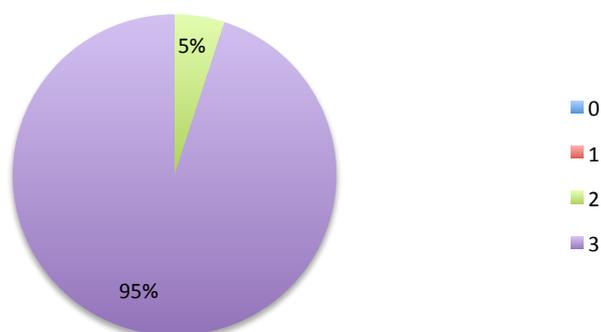


Figura B.30: Opinião sobre quantas mensagens compõe a *thread* referida anteriormente.

Spidious: Estudo com os utilizadores

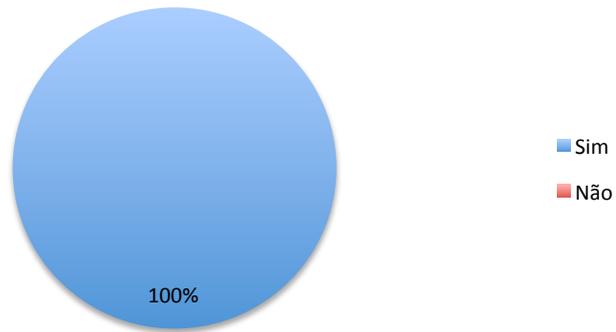


Figura B.31: Opinião sobre se existem outras *threads* na conversação referida anteriormente.

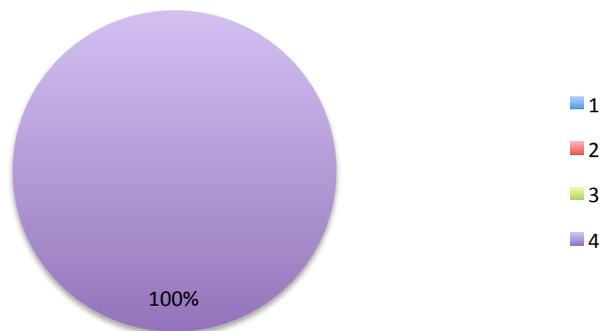


Figura B.32: Opinião sobre o número de mensagens com que a *thread*, referida anteriormente, ficou depois de efetuada uma nova resposta.

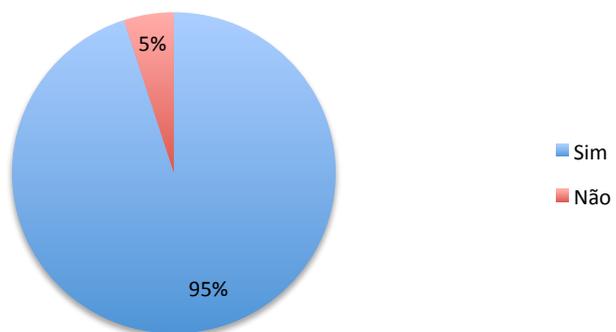


Figura B.33: Gosto pela interface de resposta rápida a uma *thread*.

Spidious: Estudo com os utilizadores

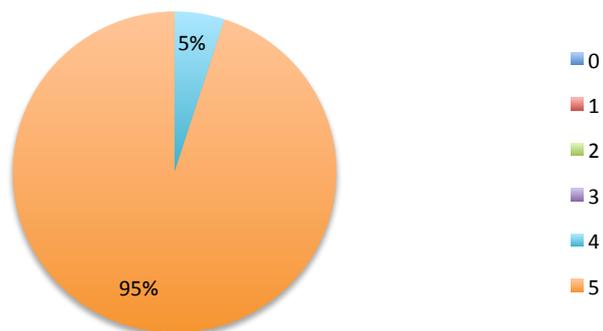


Figura B.34: Opinião sobre o número médio de palavras-chave geradas para cada *thread*.

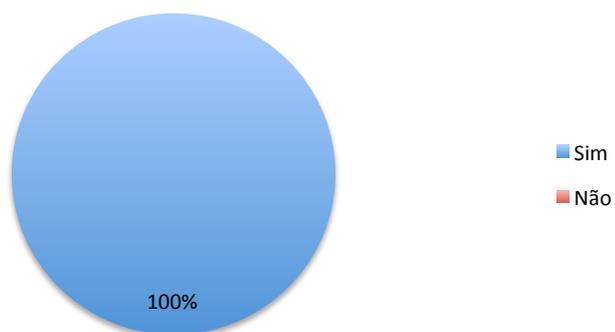


Figura B.35: Opinião sobre se as palavras-chave geradas se encontram de acordo com o conteúdo da *thread* respetiva.

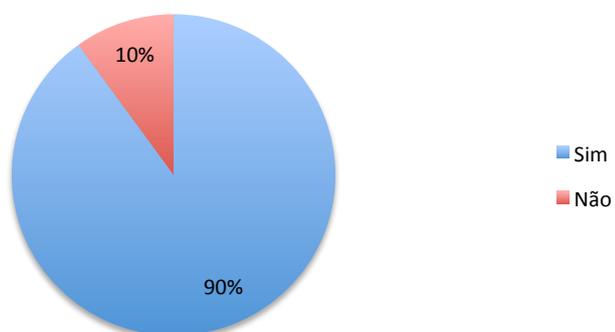


Figura B.36: Gosto pela interface da funcionalidade de palavras-chave de resumo de *thread*.

Spidious: Estudo com os utilizadores

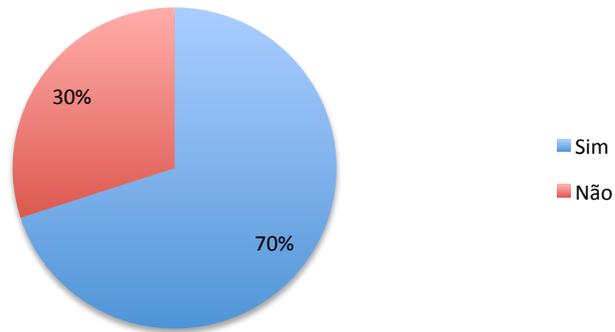


Figura B.37: Opinião sobre quão intuitiva se revelou a interface de palavras-chave de resumo de *threads*.

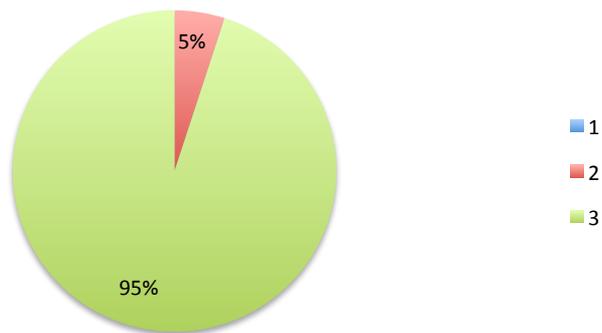


Figura B.38: Opinião sobre o número de *threads* com que a conversação com Tony Stacks ficou após a criação de uma nova *thread*.

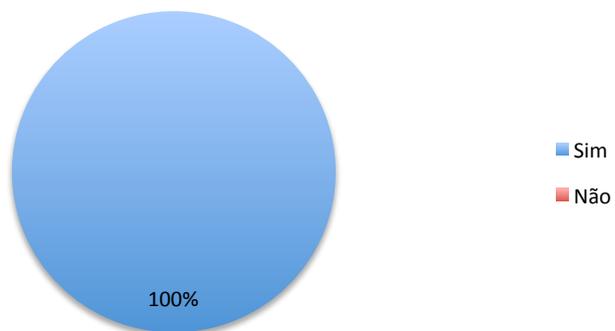


Figura B.39: Gosto pela interface implementada para a criação de uma nova *thread* com um assunto diferente, dentro de uma conversa já criada.

Spidious: Estudo com os utilizadores

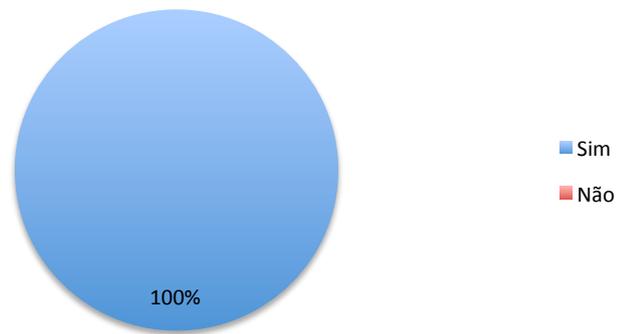


Figura B.40: Opinião sobre quão intuitivo se revelou o botão para criação de uma nova conversa.

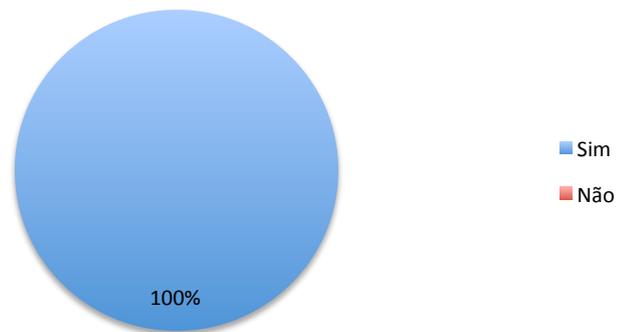


Figura B.41: Opinião sobre quão intuitiva se revelou a interface que permite efetuar a filtragem na caixa de correio eletrónico.

B.5 Pós-Questionário

Spidious - Pós-questionário do estudo com utilizadores

Prezado(a) colega, este estudo com utilizadores tem como objectivo principal aferir se as novas funcionalidades do Spidious apresentam de facto melhorias efetivas para os utilizadores, em relação aos clientes de e-mail atuais. Este pós-questionário irá essencialmente servir para perceber se o utilizador, depois de ter experimentado, sentiu realmente todas as potencialidades e vantagens esperadas com a mudança de paradigma originada pelo Spidious e por todo o seu conjunto de novas funcionalidades.

Os resultados provenientes deste estudo com utilizadores serão utilizados na minha dissertação de mestrado que tem o título: "Spidious: Improving User's Efficiency on Email with Machine Learning and a Better Visualization Approach".

Ao participar neste estudo estará certamente a contribuir para a evolução de um dos meios de comunicação digital mais usado em todo o mundo.

* Required

1. Identificar Único *

.....

Tarefas

2. Que nível de dificuldade sentiu a completar a Tarefa 1? *

Mark only one oval.

	1	2	3	4	5	
Muito Fácil	<input type="radio"/>	Muito Difícil				

3. Que nível de dificuldade sentiu a completar a Tarefa 2? *

Mark only one oval.

	1	2	3	4	5	
Muito Fácil	<input type="radio"/>	Muito Difícil				

4. Que nível de dificuldade sentiu a completar a Tarefa 3? *

Mark only one oval.

	1	2	3	4	5	
Muito Fácil	<input type="radio"/>	Muito Difícil				

5. **Que nível de dificuldade sentiu a completar a Tarefa 4? ***

Mark only one oval.

	1	2	3	4	5	
Muito Fácil	<input type="radio"/>	Muito Difícil				

6. **Que nível de dificuldade sentiu a completar a Tarefa 5? ***

Mark only one oval.

	1	2	3	4	5	
Muito Fácil	<input type="radio"/>	Muito Difícil				

7. **Que nível de dificuldade sentiu a completar a Tarefa 6? ***

Mark only one oval.

	1	2	3	4	5	
Muito Fácil	<input type="radio"/>	Muito Difícil				

8. **Que nível de dificuldade sentiu a completar a Tarefa 7? ***

Mark only one oval.

	1	2	3	4	5	
Muito Fácil	<input type="radio"/>	Muito Difícil				

Funcionalidades

9. **Como avalia a utilidade da visualização do e-mail orientada à conversa do Spidious? ***

Mark only one oval.

	1	2	3	4	5	
Muito Inútil	<input type="radio"/>	Muito Útil				

10. **Como avalia a utilidade da extração de palavras-chave para efetuar o resumo de uma thread no Spidious? ***

Mark only one oval.

	1	2	3	4	5	
Muito Inútil	<input type="radio"/>	Muito Útil				

11. Como avalia a utilidade da classificação de threads de e-mail em Trabalho e Lazer no Spidious? *

Mark only one oval.

	1	2	3	4	5	
Muito Inútil	<input type="radio"/>	Muito Útil				

12. Como avalia o resultado final da visualização do e-mail orientada à conversa do Spidious? *

Mark only one oval.

	1	2	3	4	5	
Muito Fraco	<input type="radio"/>	Muito Bom				

13. Como avalia a utilidade da filtragem do Spidious que permite, em dado momento, receber apenas e-mails de Trabalho ou e-mails de Lazer? *

Mark only one oval.

	1	2	3	4	5	
Muito Inútil	<input type="radio"/>	Muito Útil				

14. Como avalia o resultado final da extração de palavras-chave para efetuar o resumo de uma thread no Spidious? *

Mark only one oval.

	1	2	3	4	5	
Muito Fraco	<input type="radio"/>	Muito Bom				

15. Como avalia o resultado final da classificação de threads de e-mail em Trabalho e Lazer no Spidious? *

Mark only one oval.

	1	2	3	4	5	
Muito Fraco	<input type="radio"/>	Muito Bom				

16. Como avalia o resultado final da filtragem do Spidious que permite, em dado momento, receber apenas e-mails de Trabalho ou e-mails de Lazer? *

Mark only one oval.

	1	2	3	4	5	
Muito Fraco	<input type="radio"/>	Muito Bom				

17. **Quão fácil é o uso de toda a interface do Spidious, resultante da mudança de paradigma no e-mail, operada pelo mesmo? ***

Mark only one oval.

1 2 3 4 5

Muito Difícil Muito Fácil

Considerações Finais

18. **Quão melhor achou a experiência total de utilização que o Spidious lhe proporcionou, em relação à do seu cliente de e-mail atual? ***

Mark only one oval.

1 2 3 4 5

Muito Pior Muito Melhor

19. **Considera que os novos métodos de visualização do Spidious poderiam melhorar a sua eficiência de utilização no e-mail, em comparação com o seu cliente atual? ***

Mark only one oval.

Sim
 Não

20. **Considera que os algoritmos de aprendizagem do Spidious poderiam melhorar a sua eficiência de utilização no e-mail, em comparação com o seu cliente atual? ***

Mark only one oval.

Sim
 Não

21. **Considera que a conjugação da nova visualização e das técnicas de aprendizagem num cliente de e-mail web (Spidious) poderiam melhorar a sua eficiência de utilização no e-mail, em comparação com o seu cliente atual? ***

Mark only one oval.

Sim
 Não

22. **Utilizaria o Spidious no futuro caso o mesmo chegasse a fase de produção? ***

Mark only one oval.

Sim
 Não

23. Quanto estaria disposto a pagar mensalmente para obter uma subscrição do Spidious? *

Mark only one oval.

- Nada
- Entre 0 e 5 euros
- Entre 5 e 10 euros
- Mais de 10 euros

24. O que gostou mais no Spidious? *

.....

.....

.....

.....

25. O que gostou menos no Spidious?

.....

.....

.....

.....