



Tânia Ferreira Leitão

Licenciatura em Engenharia Informática

Avaliação da relevância em grupos nas aplicações de redes sociais

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática

Orientadora : Carmen Pires Morgado , Professora Auxiliar, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Co-orientador : José Alberto Cardoso e Cunha, Professor Catedrático, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Júri:

Presidente: Nuno Manuel Robalo Correia, Professor Catedrático, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Arguente: Fernando Manuel Augusto da Silva, Professor Catedrático, Faculdade de Ciências da Universidade do Porto, Departamento Ciência de Computadores

Vogal: Carmen Pires Morgado, Professora Auxiliar, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro, 2012

Avaliação da relevância em grupos nas aplicações de redes sociais

Copyright © Tânia Ferreira Leitão, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Ao meu avô João Francisco Ferreira

Agradecimentos

Desde já gostaria de agradecer à Professora Doutora Carmen Morgado e ao Professor Doutor José Cardoso e Cunha pelo o apoio, críticas e orientações para a realização desta dissertação. Também gostaria de agradecer à minha família em especial à minha mãe e pai pelo apoio e esforço que fizeram para que conseguisse acabar o meu curso. Também gostaria de agradecer a todos os meus amigos e colegas de faculdade que me apoiaram nestes últimos meses a completar o meu trabalho. Entre todos gostaria de agradecer em especial ao Filipe Carvalho, Sérgio Silva, Hélder Martins, Joana Barata, André Mourão e Ricardo Marques pelo a sua grande ajuda, motivação e troca de ideias que sempre proporcionam. Também gostaria de agradecer à empresa *Sporttools* e *Alojamento Vivo* por terem permitido usar os seus servidores para testes e por fim na versão final. Por fim gostaria de agradecer ao meu namorado Yaniel Barbosa pela sua compreensão e ajuda que nestes meses me deu.

Resumo

A agregação de indivíduos em grupos corresponde a um padrão de comportamento típico na forma como são estabelecidas as conexões nos ambientes de redes sociais, onde existe uma grande quantidade e diversidade de informação produzida.

Um das formas de conseguir seleccionar a informação que é de interesse é ter em conta a organização dos utilizadores em grupos de indivíduos interrelacionados pelas suas características, interesses comuns e interacções.

No entanto, devido à natureza dinâmica destes ambientes, existe uma necessidade constante de adaptação para que se assegure a utilidade dos grupos formados para os seus membros e se a informação partilhada se revela de utilidade para os utilizadores.

Neste trabalho propõe-se uma abordagem para avaliação da relevância da informação produzida no contexto de grupos de utilizadores em redes sociais com vista a suportar uma avaliação da utilidade dos grupos existentes. Esta abordagem contempla as seguintes vertentes: informação produzida por cada utilizador para os seus grupos; caracterização e avaliação da relevância individual de cada utilizador no contexto do grupo; análise da relevância da informação partilhada pelos seus membros no contexto de um grupo. Como forma de validar a abordagem seguida foi desenvolvida uma ferramenta num contexto específico de uma plataforma de rede social (*Facebook*), ilustrando-se assim as diferentes vertentes mencionadas.

Palavras-chave: Grupos, Relevância, Redes sociais.

Abstract

The aggregation of individuals in groups matches a typical behavioral pattern in terms of how connections are established in a social network environment, where there is a large diversity and quantity of information being produced.

One way to select useful information is to consider the user organization in groups of individuals, which are interrelated by their characteristics, common interests and interactions. Nonetheless, due to the dynamic nature of these environments, there is a constant need to adapt to ensure the utility of these formed groups.

In this work, we propose an approach to evaluate the relevance of the information being produced in the context of user groups in social networks, so we can evaluate the utility of these existing groups. This approach contemplates the following dimensions: information produced by each user within groups; characterization and evaluation of each user's relevance for the groups; analysis of the relevance of the information shared by the group members. In order to test and validate this approach a tool was developed, in a specific context of the Facebook social network platform, allowing to illustrate the above mentioned dimensions .

Keywords: Groups, Relevance, Social networks.

Conteúdo

1	Introdução	1
1.1	Motivação	2
1.2	Identificação do problema	5
1.3	Objectivos do trabalho	7
1.4	Organização da dissertação	7
2	Enquadramento e trabalho relacionado	9
2.1	Organização dos indivíduos em redes sociais	9
2.1.1	Grupos sociais	9
2.1.2	Redes sociais	10
2.1.3	<i>Sites</i> de redes sociais	10
2.1.4	Evolução dos <i>sites</i> baseados em redes sociais	11
2.1.5	Credibilidade da informação nas redes sociais	12
2.2	Extracção da informação	13
2.2.1	Perfis dos utilizadores	13
2.2.2	Filtragem, selecção e extracção de informação	15
2.2.3	Filtragem colaborativa	15
2.2.4	Análise de conteúdos	16
2.3	Grupos de utilizadores	18
2.3.1	Grupos explícitos	18
2.3.2	Grupos implícitos	18
2.3.3	Critérios de criação dos grupos	20
2.3.4	Perfis de grupos	22
2.4	Panorama dos conceitos de grupos em <i>sites</i> sociais	22
2.4.1	<i>Facebook</i>	23
2.4.2	<i>Twitter</i>	24
2.4.3	<i>LinkedIn</i>	25
2.4.4	<i>YouTube</i>	26

2.4.5	<i>Google+</i>	27
2.4.6	<i>Flickr</i>	28
2.4.7	Discussão e comparação entre os <i>sites</i> de redes sociais	28
2.5	Resumo	30
3	Modelo para a avaliação da relevância em grupos	33
3.1	Abordagem	33
3.2	Perfis de utilizador e de grupo	35
3.2.1	Perfil de utilizador	35
3.2.2	Perfil de grupo	36
3.3	Popularidade dos conteúdos	37
3.3.1	Indicadores de interesse em redes sociais	38
3.3.2	Relevância dos conteúdos	39
3.4	Relevância de utilizadores para os grupos	40
3.5	Relevância de um grupo para o utilizador	41
3.6	Idealização de uma aplicação no <i>Facebook</i>	43
3.7	Instanciação do modelo no contexto do <i>Facebook</i>	44
3.7.1	Partilha de conteúdos ou endereços - <i>Share</i>	44
3.7.2	Variável temporal	45
3.7.3	Número de conteúdos extraídos	46
3.8	Resumo	46
4	Arquitectura de suporte ao modelo	47
4.1	Organização lógica da arquitectura	47
4.1.1	Gestor de informação de grupos (<i>Groups information manager</i>)	48
4.1.2	Controlador de dados dos grupos (<i>Groups data controller</i>)	49
4.1.3	Gestor de informação de utilizadores (<i>User information manager</i>)	49
4.1.4	Controlador de dados do utilizador (<i>User data controller</i>)	49
4.1.5	Controlador de correspondências (<i>Match controller</i>)	50
4.1.6	Controlador de relevância de um utilizador nos grupos $U \rightarrow G$ (<i>User Group relevance controller</i>)	50
4.1.7	Controlador de relevância de um grupo para o utilizador $G \rightarrow U$ (<i>Group User relevance controller</i>)	50
4.1.8	Controlador de estatísticas (<i>Statistics center</i>)	50
4.1.9	Bravo Boss (<i>Bravo controller</i>)	51
4.1.10	Controlador da base de dados (<i>Database controller</i>)	51
4.2	Dados obtidos através do <i>Facebook</i>	51
4.2.1	Perfil do utilizador	51
4.2.2	Perfil de grupo	52
4.3	Dados obtidos através da <i>API</i> do <i>Alchemy</i>	54
4.4	Aplicações no <i>Facebook</i>	55

4.5	Resumo	56
5	Implementação	57
5.1	Camada de base de dados	58
5.1.1	Estrutura da base de dados	59
5.1.2	Desenho da base de dados	59
5.1.3	Informação dos grupos na base de dados	60
5.2	Camada de sistema	61
5.2.1	Camada de modelos	62
5.2.2	Camada de controladores	67
5.3	Camada de aplicação	77
5.4	Camada de apresentação	78
5.4.1	Estatísticas (<i>Stats</i>)	79
5.4.2	Relevância dos grupos (<i>Groups relevance</i>)	79
5.4.3	Relevância dos utilizadores (<i>Users relevance</i>)	81
5.4.4	Partilha de informação (<i>Share Information</i>)	81
5.5	Resumo	81
6	Avaliação experimental	83
6.1	Descrição do ambiente e universo de utilizadores	83
6.2	Resultados experimentais	85
6.2.1	Comportamento dos utilizadores nos grupos	85
6.2.2	Relevância de utilizadores para os grupos	87
6.2.3	Importância dos grupos para os utilizadores	90
6.2.4	Correspondência entre conteúdos e grupos	92
6.3	Inquéritos de satisfação e análise de resultados	93
6.3.1	Inquérito de satisfação	94
6.3.2	Tipos de utilizadores	95
6.3.3	Relevância de utilizadores	95
6.3.4	Relevância de grupos	96
6.3.5	Partilha de conteúdos para os grupos	98
6.3.6	Sugestões da parte dos utilizadores	99
6.4	Resumo	100
7	Conclusões e trabalho futuro	101
7.1	Conclusões	101
7.2	Trabalho futuro	102

Lista de Figuras

1.1	Distribuição de utilizadores na Internet - [retirado <i>CIA's World Factbook</i>]	2
1.2	Distribuição de utilizadores nos <i>sites</i> de redes sociais - [retirado de <i>Pew Internet & American Life Project</i>]	3
2.1	Representação de uma rede social	10
2.2	Divisão dos participantes pelos <i>sites</i> de redes sociais no Mundo em Dezembro de 2011- Retirado de <i>Google Trends</i> (Créditos : Vincenzo Cosenza)	12
2.3	OpenCalais	17
3.1	Abordagem geral do modelo	34
3.2	Diagrama do perfil de utilizador	36
3.3	Perfil grupo	37
4.1	Arquitectura da aplicação no <i>Facebook</i>	48
4.2	Perfil de grupo no <i>Facebook</i>	53
4.3	Janela do <i>Facebook</i> a pedir autorização para a instalação da aplicação.	56
5.1	Arquitectura interna da aplicação <i>GIA</i>	58
5.2	Arquitectura do Model-View-Control	59
5.3	Estrutura da base de dados	60
5.4	Relações entre as tabelas	61
5.5	Gráficos da relevância de um utilizador no grupo	78
5.6	Imagem do processamento da aplicação <i>GIA</i>	79
5.7	Janela da aplicação <i>GIA</i> com as estatísticas dos grupos	80
5.8	Janela da aplicação <i>GIA</i> com a relevância dos grupos para um utilizador	80
5.9	Janela da aplicação <i>GIA</i> com as importâncias dos utilizadores nos grupos	82
5.10	Janela da aplicação <i>GIA</i> com a partilha de informação nos grupos	82
6.1	Estatísticas do <i>Facebook</i> da aplicação <i>GIA</i>	84
6.2	Distribuição de grupos analisados no <i>GIA</i>	85

6.3	Participação de utilizadores em vários grupos específicos I	87
6.4	Participação de utilizadores em vários grupos específicos II	88
6.5	Participação em vários tipos de grupos I	89
6.6	Participação em vários tipos de grupos II	90
6.7	Grupo <i>INOV</i> analisado através do <i>GIA</i>	91
6.8	Grupo <i>Tasqueiros de Dresden</i> analisado através do <i>GIA</i>	91
6.9	Grupo <i>Jantar de Natal 2011</i> analisado através do <i>GIA</i>	92
6.10	Relevância de grupos para o utilizador João	92
6.11	Relevância de grupos para a utilizadora Maria	93
6.12	Utilização de redes sociais	95
6.13	Interesse de calcular a popularidade de utilizadores	96
6.14	Interesse de uma ferramenta que verifique em que grupos um utilizador tem interesse	96
6.15	Interesse de uma ferramenta que verifique que grupos um utilizador tem interesse	97
6.16	Popularidade dos grupos para o utilizador na aplicação <i>GIA</i>	97
6.17	Interesse numa funcionalidade de partilha de informação para os grupos consoante o tema do grupo	98
6.18	Avaliação da partilha de informação na aplicação <i>GIA</i>	99

Lista de Tabelas

1.1	Internet por continente -[retirado de <i>World Internet Usage Statistics News and World Population Stats</i>]	2
2.1	Operações disponibilizadas pelo grupo no <i>Facebook</i>	23
2.2	Operações disponibilizadas pelo <i>Twitter</i>	25
2.3	Operações nos grupos disponibilizadas pelo <i>Linkedin</i>	26
2.4	Operações disponibilizadas pelo <i>Youtube</i>	26
2.5	CrITÉrios para a sugestão de novos utilizadores no <i>Google+</i>	27
2.6	Operações sobre os grupos disponibilizadas pelo <i>Flickr</i>	28
2.7	Operações disponibilizadas pelas <i>APIs</i> a nível da informação base	29
2.8	Operações disponibilizadas pelas <i>APIs</i> a nível da informação estendida - (I)	29
2.9	Operações disponibilizadas pelas <i>APIs</i> a nível da informação estendida - (II)	30
2.10	Operações disponibilizadas pelas <i>APIs</i> a nível da informação estendida - (III)	30
3.1	Indicadores de interesse - Variáveis	38
3.2	Média dos indicadores de interesse nos grupos	39
3.3	Componentes da relevância de grupos para o utilizador abrangendo todos os conteúdos partilhados	42
3.4	Propriedades da relevância do grupo para o utilizador	43
4.1	Comandos da <i>API</i> para aceder à informação sobre os <i>posts</i> do utilizador .	52
4.2	Perfil de grupo no <i>Facebook</i>	52
4.3	Perfil de grupo estendido no <i>Facebook</i>	53
5.1	Métodos do Modelo Correspondências	63
5.2	Métodos do Modelo Utilizador	63
5.3	Métodos do Modelo Estatísticas (<i>Stats</i>)	64
5.4	Métodos do Modelo Estatísticas Utilizador Grupo(<i>StatsUserGroup</i>)	64
5.5	Métodos do Modelo Grupo	65

5.6	Métodos do Modelo Estatísticas Grupo Utilizador(<i>StatGroupUser</i>)	66
5.7	Métodos do Modelo Sessão	66
5.8	Métodos do Modelo Sessão Grupo	67
5.9	Constantes na expressão do cálculo da relevância de utilizadores	73
5.10	Constantes na expressão do cálculo da relevância de grupos para o utilizador	73
5.11	Legenda para os valores da relevância dos grupos para o utilizador	81
6.1	Participação nos diferentes tipos de grupos no <i>Facebook</i>	86
6.2	Participação nos diferentes tipos de grupos no <i>Facebook</i> II	86

Listings

5.1	Inicialização das configurações	68
5.2	Autenticação do utilizador	69
5.3	Cálculo da popularidade de um <i>post</i>	70
5.4	Método auxiliar para o cálculo da popularidade de um <i>post</i>	71
5.5	Contagem dos <i>likes</i> dos posts	71
5.6	Calculo do valor temporal	72
5.7	Chamada ao servidor do <i>Alchemy</i>	75
5.8	Resposta em formato <i>JSON</i> do servidor do <i>Alchemy</i>	76



Introdução

À medida que a Internet vai crescendo a nível mundial, a quantidade de informação que se vai aglomerando torna a procura e gestão da informação cada vez mais difíceis. A Internet tem vindo a tornar-se cada vez mais importante nas actividades do quotidiano das pessoas, pois permite que possam comunicar e relacionar-se mutuamente com maior facilidade, sem inconveniência da distância e de uma forma livre [Que10]. Entre 2000 e 2010 o número de utilizadores na Internet, subiu cerca de 444,8 %, como podemos observar na Tabela 1.1 e na Figura 1.1, em grande parte devido à grande quantidade de serviços disponíveis, tais como troca de mensagens electrónicas, serviços de acesso remoto, colaboração ou partilha de ficheiros entre utilizadores. Este novo meio de comunicação permitiu que fossem ultrapassadas barreiras como grandes distâncias ou fronteiras que tornavam a comunicação, relacionamento e divulgação da informação tão difíceis [Enz01]. Esta evolução na Internet, motivou o aparecimento e crescimento de novos tipos de aplicações vocacionadas para diversas áreas. As novas aplicações sociais, mudaram completamente a vida das pessoas em geral e facilitaram muito a comunicação e o relacionamento, mudando hábitos do quotidiano.

A insegurança e a vida sedentária verificada na sociedade, leva a que muitas pessoas passem a maior parte do seu tempo em sua casa no seu computador. As mudanças provocadas vieram dar a possibilidade aos utilizadores de interagirem, pesquisarem e partilharem informação com outras pessoas, empresas e entidades, sem saírem de suas casas.

O aparecimento e a utilização dos novos tipos de aplicações e serviços na Internet, tais como os *sites* colaborativos, *blogues* e *sites* de redes sociais vieram evidenciar a importância destas aplicações interactivas. No caso das aplicações baseadas em redes sociais, uma pessoa pode fazer parte destas registando um utilizador, pelo qual é identificada. Este

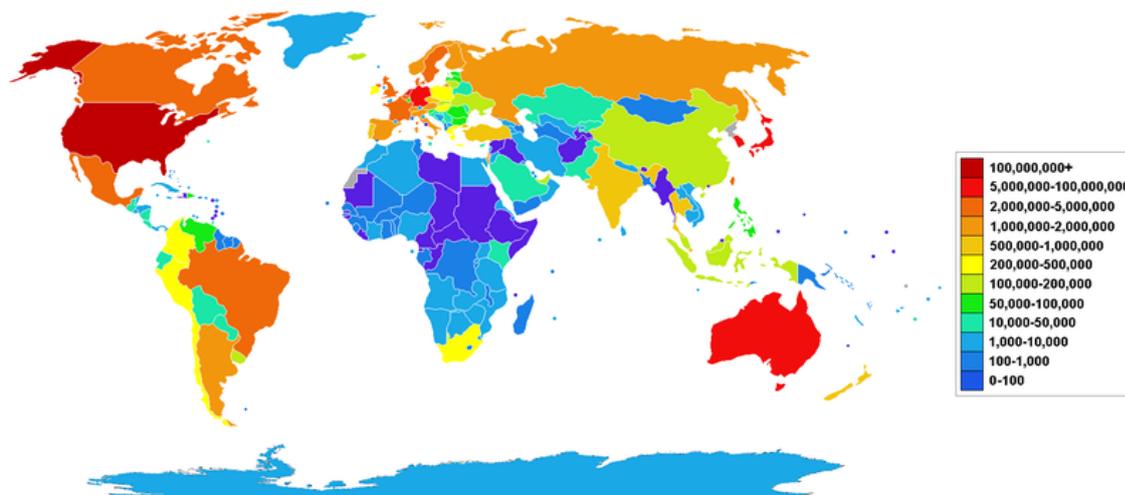


Figura 1.1: Distribuição de utilizadores na Internet - [retirado CIA's World Factbook]

Continentes	População (2010)	Utilizadores de Internet (2000)	Utilizadores de Internet(2010)	Crescimento
África	1.013.779.050	4.514.400	110.931.700	2.357,3
América do Norte	344.124.450	108.096.800	266.224.500	146,3
América Latina/Caribe	592.556.972	18.068.919	204.689.836	1.032,8
Ásia	3.834.792.852	114.304.000	825.094.396	621,8
Médio Oriente	212.336.924	3.284.800	63.240.946	1.825,3
Europa	813.319.511	105.096.093	475.069.448	352,0
Oceania	34.700.201	7.620.480	21.263.990	179,0
Total	6.845.609.960	360.985.492	1.966.514.816	444,8

Tabela 1.1: Internet por continente -[retirado de World Internet Usage Statistics News and World Population Stats]

utilizador por norma tem um perfil associado, onde pode colocar a informação sobre si e sobre os seus gostos e preferências consoante o conceito da aplicação. É assim possível expor informação sobre os utilizadores, tal como a sua identidade, interesses, crenças políticas e religiosas, etc. Estas informações, disponibilizadas nos perfis de utilizador, podem ser estudadas pelos analistas das aplicações e serem utilizadas de forma a tentar identificar consoante a relevância da informação, formas de filtrar apenas a informação de interesse para o utilizador.

1.1 Motivação

Com o crescimento da Internet, as pessoas começaram a desenvolver redes sociais e a partilhar bastantes informações que podem ser acedidas através das mesmas. Desde 2005 o crescimento dos utilizadores nos *sites* de redes sociais tem vindo a aumentar significativamente [Pro09]. Estudos realizados nos Estados Unidos da América mostram que 83% da população adulta entre os 18 e os 29 anos, utiliza os *sites* de redes sociais como se pode verificar na Figura 1.2. Hoje em dia, as pessoas relacionam a sua vida física com a sua realidade na rede, tornando-as integradas. As interações físicas tornaram-se também virtuais o que permitiu a construção automática das redes sociais e da sua organização

social [Cas00].

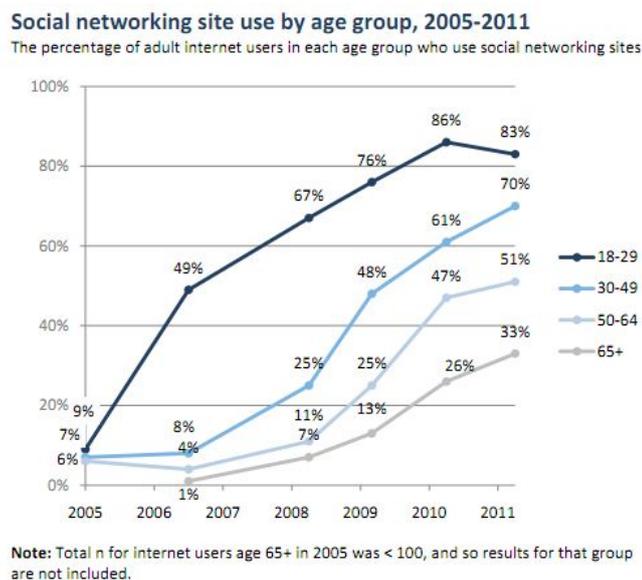


Figura 1.2: Distribuição de utilizadores nos *sites* de redes sociais - [retirado de *Pew Internet & American Life Project*]

O crescimento destas aplicações sociais deveu-se em parte ao facto da sua utilização ser realizada com baixos custos, de forma segura e praticamente de forma instantânea. Outra razão surge do facto de estes *sites* permitirem que os utilizadores conheçam novas pessoas, retomem relações antigas e, mais do que isso, permite que um utilizador torne visível a sua rede social ou por outras palavras, a sua rede de contactos. A divulgação pública da rede social entre utilizadores é um componente de interesse crucial pois permite que se possa verificar quem pertence à sua rede e visualizar os perfis de cada utilizador dependendo das suas políticas de privacidade. Desta forma, através de um utilizador tem-se uma vasta rede de pesquisa de contactos. A divulgação, ou não, de características pessoais de um utilizador revela o que este quer representar no seu perfil no *site* social [Gof59]. Outro motivo para o grande número de utilizadores destes *sites* de redes sociais, deve-se ao facto de bastantes pessoas gostarem de aderir a inovações tecnológicas que facilitem as interacções sociais. Os *sites* de redes sociais solucionaram uma das necessidades da sociedade que ainda não estava respondida pela tecnologia, a de comunicar de forma livre e espontânea, advindo daí a sua popularidade [Que10]. O facto desta forma de comunicar ser um meio bastante fácil de propagar a informação, também tornou os *sites* de redes sociais bastante interessantes já que, através de partilhas de informação no seu perfil ou no de outros utilizadores, se pode partilhar informação que, por outros meios de comunicação, poderia ser processada lentamente e ter associados custos mais elevados. O interesse num *site* de rede social está também muito associado às pessoas manterem um interesse sobre os contactos com quem se relacionam e o que estes fazem, partilham ou comentam. Desta forma um indivíduo é caracterizado pela informação partilhada, pelo seu conjunto específico de características, pelas suas interacções e pela sua

rede social de contactos.

Existem diversos tipos de sistemas e aplicações que tiram partido da informação e dados dos utilizadores. Uma forma de aproveitar os dados sobre as características individuais e as interacções entre utilizadores através das redes sociais dos utilizadores, é ilustrada nos sistemas de recomendação [JS07]. Os sistemas de recomendação tentam encontrar informações que lhes permitem, com base em certas métricas, devolver sugestões ao utilizador como o que comprar, o que ouvir ou quem adicionar como amigo. Existem sistemas de recomendação que através de estimativas de vários utilizadores calculam e decidem que escolhas a fazer. Como exemplo, o caso do sistema *Flytrap* [CBH02] que escolhe que músicas devem ser tocadas num espaço, através de estimativas dos gostos dos utilizadores presentes nesse espaço. Estas recomendações devem criar e gerir sugestões, de forma a abranger um grande número de soluções aos utilizadores. Actualmente muitos dos sites das redes sociais que existem, como o *Facebook* [Fac11a] ou *Google+* [Goo11a], implementam alguma funcionalidade relacionada com recomendação.

Porém, o conceito de rede e grupo social não tem ficado restringido aos *sites* de redes sociais. Na área dos jogos desde logo surgiram redes sociais criadas através da interacção entre jogadores. Nestes jogos virtuais como o *World of Warcraft*¹ ou *Call of duty*², não é exigido ao utilizador que divulgue qualquer informação sobre si para poder jogar. No entanto as interacções directas entre utilizadores acontecem de forma a criarem grupos para poderem mais facilmente completar os objectivos do jogo. Com características diferentes existem as redes de jogos sociais como é o caso do *Second Life*³, que tenta replicar todas as dimensões das interacções efectuadas por uma pessoa frente a frente, criando uma rede social e interacções que apenas são virtuais.

A adesão às redes sociais não ficou somente pelos indivíduos mas também chegou às próprias empresas que viram nas redes sociais uma oportunidade de negócio. As empresas tiveram interesse em aderir, com o objectivo de poderem publicitar e vender novos produtos, angariar novas parcerias e discutir e resolver assuntos de cariz profissional. Estas começaram a tentar tirar partido da súbita popularidade destas comunidades como forma de publicitar os seus produtos de forma mais fácil por forma a poderem adquirir opiniões dos consumidores antes e após o lançamento dos produtos. O facto de um produto poder ser criticado de forma positiva pelos clientes, pode influenciar de forma decisiva o sucesso do mesmo. A título de exemplo num quadro de um estudo da *Wetpaint* e *Altimeter Group* [Alt12] observou-se que as empresas que aderiram às redes sociais

¹*World of Warcraft*- É um jogo *online* MMORPG (Massively Multiplayer Online Role-Playing Game) da produtora *Blizzard* de acção e aventura [Bli12].

²*Call of Duty* -É um jogo de simulação de guerra na primeira pessoa bastante popular [Act12].

³*Second Life*- É um jogo que decorre num ambiente virtual que simula em alguns aspectos a vida real e social de um ser humano [Sec12].

aumentaram os lucros em cerca de 18% [Sch09]. Estas redes também permitem a colaboração interna entre funcionários, de forma a manterem-se actualizados e de forma a poderem expressar as suas opiniões gerais. Não só as redes sociais mas também as novas aplicações colaborativas mudaram o conceito de trabalho em grupo, vindo assim permitir o aumento da rapidez com que a informação é partilhada, não sendo necessário que os colaboradores estejam no mesmo espaço físico. Através dos *sites* de redes sociais, as empresas poderão também fazer pesquisas centradas (ou baseadas) nos dados, para perceber quem poderão ser os seus potenciais clientes, colaboradores e parceiros.

Hoje em dia existem várias empresas e entidades públicas que usam as redes sociais para poderem publicitar e divulgar informação sobre os seus produtos e serviços. Por exemplo, algumas empresas usam as redes sociais para lançamentos de novos produtos como o jogo *Tomb Raider*⁴ e alguns políticos fazem comunicados à população como o Presidente da República Portuguesa, Aníbal Cavaco Silva [Sil11], através do *Facebook*.

1.2 Identificação do problema

Como já foi referido, existe uma grande quantidade de informação que hoje em dia se encontra presente e acessível através da Internet, o que fez surgir a necessidade de existirem formas de a organizar e filtrar. Uma das formas de organizar e seleccionar a informação na qual haja interesse, é através da formação de grupos de utilizadores que são definidos pelas suas características colectivas, tais como os seus interesses comuns e a sua integração em comunidades de utilizadores. Através da análise da informação associada aos perfis dos utilizadores, é possível definir conjuntos de indivíduos interrelacionados pelas suas características, interesses em comum e interações e dessa maneira formar grupos. No entanto, nem sempre se observa que esses grupos possam ter alguma utilidade para o utilizador, ou mesmo que este seja importante para o grupo. Na base da formação de grupos podem ser consideradas as características individuais dos utilizadores, tais como a idade, cidade onde vive ou a própria nacionalidade. Também os dados relacionados com os interesses manifestados pelo utilizador, podem ser considerados, quer de forma explícita - por exemplo, relativamente a um determinado tema que associou ao seu perfil - ou quer de forma implícita, identificando interesses através do comportamento dinâmico do utilizador - por exemplo, ao consultar documentos ou visitar páginas *Web*. Trabalhos anteriores relacionados com estes conceitos de grupos mostram a utilidade do conceito para o desenvolvimento de aplicações e conduziram, em particular, ao desenvolvimento de uma ferramenta *TIG (Tool for Implicit Groups)* [Pai11], vocacionada para a gestão de grupos criados de forma implícita e que foi aplicada no caso do ambiente colaborativo *Moodle* [Moo12].

No entanto, nesses trabalhos não se aprofundou o estudo de uma dimensão particularmente crítica para o desenvolvimento de aplicações baseadas no conceito de grupos,

⁴Tomb Raider - Tomb Raider é uma série de jogos e filmes tendo como protagonista a personagem Lara Croft [Rai12].

relacionada com o problema da caracterização ou identificação da relevância destes grupos e da informação por eles produzida para os utilizadores. Ao contrário dos modelos baseados em grupos explícitos, em que é o próprio utilizador ou gestor da aplicação que assegura a relevância da informação utilizada para efeito da formação ou manutenção dos grupos, esta relevância torna-se muito difícil de identificar e avaliar quando se trata de grupos formados ou geridos implicitamente, pois os utilizadores podem partilhar diversas informações ou manifestarem determinados interesses pelas suas acções, que na verdade podem nem ser importantes em determinados contextos de aplicação. É, pois, importante que o sistema descubra por si, qual a informação que deve ter em conta e qual a que deve ser ignorada. Para isso torna-se necessário identificar o conjunto de dimensões que permitem avaliar aquela relevância, em cada contexto específico de utilização. E, se é verdade que não é possível conseguir uma solução completamente genérica para este problema, independentemente do contexto de aplicação, compreende-se que a identificação do quadro global daquelas dimensões possa ajudar a resolver o problema em cada caso. A identificação da relevância neste problema tem duas dimensões diferentes a considerar.

Uma das dimensões relaciona-se com o problema de determinar qual a relevância da informação encontrada num contexto de grupos de utilizadores. Existe informação que é partilhada por utilizadores nos grupos, de uma forma pouco exacta, provocando muitas vezes partilhas de informações que não estão envolvidas no tema do grupo. Dessa forma é importante identificar a relevância que os utilizadores têm para um determinado grupo baseado no conteúdo que partilham e perceber quem são utilizadores que realmente contribuem para os interesses daquele grupo. O facto de se poder extrair das informações, tópicos que sugerem os temas de cada grupo, permite analisar de entre as informações que são partilhadas, quais as que são realmente do âmbito daquele grupo. Isso é dependente de quão relevante é a informação que é extraída pelas ferramentas de análise e extracção de conteúdos. Por exemplo, se uma informação partilhada sobre Física for disseminada em um grupo caracterizado pelo tópico "línguas estrangeiras", provavelmente essa informação deverá ser ignorada pois foge do âmbito do grupo e provavelmente a sua relevância será nula.

A outra dimensão situa-se a jusante da primeira, relacionando-se com o problema da avaliação da relevância dos grupos para o próprio utilizador, ou seja, calcular a relevância de um grupo face ao próprio utilizador. Se um utilizador, por exemplo, deixar de consumir informação de um determinado grupo, então isso poderá indicar que esse grupo poderá não ser relevante ao mesmo. Estes conceitos e dimensões foram desenvolvidos neste trabalho, em que se estudou o conceito da relevância da informação no caso particular da disseminação de informação para os grupos e a relevância dos próprios grupos para o utilizador em relação à continuidade e utilidade da sua participação.

Na perspectiva do utilizador, é importante que o sistema seja capaz de regular de forma eficaz o processo de gestão dos grupos, com o objectivo de controlar a explosão descontrolada de grupos bem como evitando a manutenção de grupos que se tenham

tornado progressivamente irrelevantes.

É também importante referir que, na perspectiva do sistema ou ferramenta de análise da relevância de grupos, os cálculos para realizar aquela avaliação devem ser efectuados de forma ponderada para não sobrecarregar as aplicações com informação pouco relevante. Visto isto, é importante classificar e analisar os conteúdos publicados e interacções realizadas, de forma a que se dê valor à informação que realmente é importante para o utilizador e para o grupo, excluindo assim informação que seja pouco relevante ao sistema, de uma forma ponderada, sem que seja necessário sobrecarregar o sistema com grupos desnecessários ou informação pouco relevante.

1.3 Objectivos do trabalho

Devido aos problemas enumerados na secção anterior, consideramos interessante o desenvolvimento de ferramentas que dêem suporte à categorização da relevância da informação produzida pelos utilizadores. Dessa forma colocou-se, como objectivo deste trabalho, a concepção e desenvolvimento de uma ferramenta que, tanto através dos perfis pessoais dos utilizadores, como através dos perfis dos próprios grupos em uma determinada rede social, irá gerir a informação de forma a que seja possível calcular e avaliar a relevância da informação em três diferentes vertentes:

1. Disseminação selectiva da informação pelos diferentes grupos a que um utilizador pertence. Esta disseminação é baseada nos conteúdos da informação produzidos e na sua adequação aos perfis dos grupos;
2. Caracterização da informação produzida pelo utilizador no contexto dos grupos, para avaliar a sua utilidade (relevância), para os membros do grupo. Esta avaliação evolui através da participação colectiva dos restantes membros do grupo que avaliam a relevância da informação produzida por cada utilizador, conduzindo a uma medida de "popularidade" associada a um membro de um grupo;
3. Caracterização da relevância da informação partilhada no contexto de grupo para os membros do grupo. Esta avaliação permite aferir a importância de cada grupo para o utilizador, através da relevância da informação partilhada e mantida pelo grupo. Isto permite também inferir uma medida de "popularidade" associada à informação partilhada pelo grupo.

Considerou-se também, como objectivo deste trabalho, a validação da aplicabilidade da própria ferramenta através da realização de vários testes com utilizadores de uma plataforma de rede social, *Facebook*.

1.4 Organização da dissertação

Este documento está dividido em sete capítulos e uma secção bibliográfica onde estão todas as referências utilizadas neste documento.

O primeiro capítulo (Capítulo 1), "Introdução", introduz o tema do trabalho, identificando o problema e a abordagem planeada para a sua resolução.

O segundo capítulo (Capítulo 2), intitulado "Enquadramento e trabalho relacionado", descreve os conceitos relacionados com o tema abordado em torno dos grupos, fazendo uma breve consolidação de algumas definições necessárias para a leitura do documento. Apresentam-se também temas relacionados com as características dos utilizadores e dos grupos e descrevem-se alguns trabalhos realizados nessa área. Após essa reflexão abordam-se as várias formas de criação de grupos e refere-se o trabalho existente nesta área específica. Ainda neste capítulo, apresenta-se uma contextualização das várias aplicações sociais e qual a sua relevância para a resolução do problema.

No terceiro capítulo (Capítulo 3), "Modelo para a avaliação da relevância em grupos", apresenta-se uma primeira perspectiva sobre as dimensões fundamentais para seleccionar o problema através da contabilização da relevância do utilizador, dos próprios conteúdos e do grupo. Apresenta-se um modelo para definir a relevância de conteúdos através de factores, resultantes da observação empírica de comportamentos em redes sociais. Ainda neste capítulo, existe uma instanciação do problema ao caso real e à plataforma utilizada, o *Facebook*, onde se descrevem as adaptações necessárias a serem realizadas ao modelo de avaliação.

No quarto capítulo (Capítulo 4), "Arquitectura de suporte ao modelo", propõe-se uma arquitectura que suporta o modelo proposto no terceiro capítulo. Descrevem-se também, os principais componentes arquitectónicos e os fluxos de tarefas propostos para a aplicação que executará através de uma aplicação no *Facebook*.

No quinto capítulo (Capítulo 5), "Implementação", existe uma descrição pormenorizada da implementação do cálculo das relevâncias e das várias camadas que a compõem. É descrita a forma como a implementação foi efectuada, não só as linguagens mas também quais os problemas e decisões que foram tomadas na fase de implementação. Também é descrita a forma como esta aplicação funciona e uma demonstração da própria interface.

No sexto capítulo (Capítulo 6), "Avaliação experimental", existe uma descrição de todos os resultados obtidos, incluindo os inquéritos, realizados a um conjunto de utilizadores que testaram a aplicação proposta. Acompanhando a mostra de resultados, fazem-se alguns comentários e conclusões sobre os dados obtidos.

Por fim no sétimo capítulo (Capítulo 7), "Conclusões e trabalho futuro", são apresentadas as conclusões resultantes da elaboração desta dissertação e apresentadas algumas ideias para trabalho futuro, para poder prosseguir as ideias propostas desta dissertação.



Enquadramento e trabalho relacionado

Neste capítulo apresenta-se uma análise de conceitos relacionados com a formação e definição de grupos, perfis de utilizador e de grupos, bem como uma caracterização das aplicações de redes sociais, focada essencialmente nas funcionalidades oferecidas ao nível dos grupos.

2.1 Organização dos indivíduos em redes sociais

Desde a antiguidade que as pessoas se agrupavam para conseguir alcançar objectivos comuns, em actividades como a caça, a construção e o comércio [NRW12]. Cedo se percebeu que estes grupos contribuíram para desenvolver o grau de socialização permitindo desenvolver aptidões da civilização como as conhecemos hoje em dia. Com a evolução da civilização, os grupos sociais tornaram-se cada vez mais relevantes para as relações interpessoais e para a própria sobrevivência.

2.1.1 Grupos sociais

Em sociologia um grupo social é um conjunto de pessoas que interagem sucessivamente formando uma colectividade bem identificada e organizada [Boy08]. Um grupo é um conjunto estruturado de pessoas ou objectos com características ou objectivos em comum que se agrupam numa entidade, por possuírem aspectos e interesses mútuos [Pre12]. Quando um grupo está formado são construídas propriedades através das suas características como a sua natureza, número de elementos ou mesmo o tempo de vida do próprio

grupo. Um grupo irá ser visto como um conjunto de pessoas que compartilham certas características e interesses, interagem entre si, aceitam direitos e obrigações como pertencentes ao grupo e compartilham uma identidade comum. Os interesses que podem levar à criação de um grupo podem situar-se nas mais diferentes áreas como educação, negócios, amizade, amor ou mesmo por distintas ideologias. As interações entre indivíduos e mesmo entre os grupos criaram redes sociais que se foram desenvolvendo, como podemos ver na Subsecção 2.1.2.

2.1.2 Redes sociais

Uma rede social é uma estrutura composta por pessoas ou organizações que podem ser vistas como nós da rede, que estão ligados através de arcos, segundo os seus relacionamentos de interesses em comum, amizade, financeiro, sexual ou crenças. Uma rede social pode ser representada sob a forma de um grafo - designado como grafo social como se ilustra na Figura 2.1. A rede social de um indivíduo pode ter um número de indivíduos variável sendo que a estrutura e o tamanho podem ser bastante amplos e ter vários graus de ligação. A força da ligações entre utilizadores é mais forte ou mais fraca dependendo do tipo de interacção, por exemplo ser mais frequente ou menos frequente. Ao caracterizar a raiz do grafo como correspondente ao nó que identifica um dado indivíduo, define-se a rede social do indivíduo com as suas correspondentes forças de ligação (Figura 2.1) representadas através da espessura do arco, onde o arco mais espesso representa uma ligação mais forte. Através destes relacionamentos vão-se criando grupos de indivíduos que têm forças de ligação mais elevadas tornando assim interessante a formação de grupos baseados na interacção [RBDD⁺10].

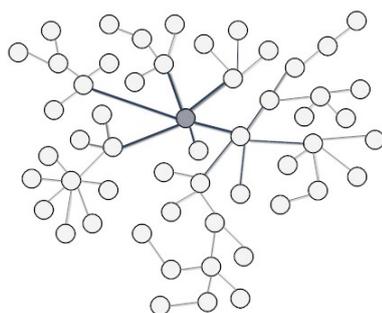


Figura 2.1: Representação de uma rede social

2.1.3 Sites de redes sociais

Os *sites* de redes sociais alojam ou suportam serviços baseados na Internet, associados a redes de utilizadores, que permitem que os indivíduos construam perfis públicos ou semi-públicos com informações sobre estes e ainda criem listas de ligações com outros utilizadores com quem se relacionam [BE07]. Os utilizadores podem dessa forma interagir, partilhar e comunicar de forma pública ou privada informação muito diversificada,

desde fotografias, vídeos ou mensagens. Um utilizador pode definir as suas características e interesses no seu perfil pessoal que está associado à sua identidade naquela rede social. O que tornou estes *sites* de redes sociais tão únicos foi a sua capacidade de capturarem a dinâmica das interações pessoais, não só permitindo a comunicação com indivíduos que não conhecemos como também a divulgação e visualização da nossa rede social de conhecidos.

Estas redes podem ser divididas em diferentes categorias e âmbitos, como redes de relacionamentos informais de conteúdo pessoal como o *Facebook* e *Twitter* [Twi11a], redes com objectivos profissionais como o *LinkedIn* [Lin12a] ou redes sociais a nível dos jogos virtuais como *World of Warcraft* [Bli12] e jogos sociais como o *Second Life* [Sec12].

2.1.4 Evolução dos *sites* baseados em redes sociais

Os primeiros *sites* baseados em redes sociais começaram a surgir a partir de 1997, sendo de destacar uma ferramenta chamada *SixDegrees*¹ que permitia aos utilizadores criar perfis, enviar mensagens, listar amigos e navegar nas redes sociais dos próprios. Apesar da adesão de inúmeros utilizadores, o *SixDegrees* não conseguiu a sustentação financeira, o que resultou na interrupção do serviço mais tarde. Pode-se dizer que esta ferramenta surgiu antes do tempo, visto que os utilizadores ainda não estavam preparados para este novo conceito e modelo de negócio. A partir de 1997, surgiram diversos novos *sites* com diversas combinações de tipos de perfis e de publicações, como o *LiveJournal* [Fit99], *Friendster*² e *FotoLog*³.

Os *blogues*⁴ como o *LiveJournal* permitiam que os utilizadores pudessem manter um diário *online* na sua comunidade virtual. A maior parte das novas aplicações eram tentativas de cópias das anteriores com algum tipo de especialização. O *site* que mais se assemelhava aos *sites* de redes sociais como agora são conhecidos foi o *Friendster*, mas o grande sucesso levou ao seu colapso e como consequência, acabou por limitar as funcionalidades aos seus utilizadores em função da crescente demanda [BE07].

A competir com o *Friendster* surgiu o *MySpace*⁵ que queria aliciar utilizadores do *Friendster* a migrarem para o seu serviço. Com o passar do tempo, o *MySpace* especializou-se como um *site* focado na área da Arte, principalmente na música. Neste *site* várias bandas partilhavam notícias que podiam ser vistas e comentadas, bem como músicas que os seus *fans* poderiam ouvir. Neste *site* também foram adicionadas novas funcionalidades de acordo com as especificidades suas próprias. A popularidade de cada *site* depende muito da região pois existem certos *sites* com bastante sucesso na América mas que na

¹*SixDegrees*- Foi uma rede social criada em 1997 que permitia listar amigos, família e conhecidos no próprio *site* ou em outros [Six12].

²*Friendster* -Era uma rede social fundada em 2002 por Jonathan Abrams em Mountain View, California. Actualmente é uma rede social de jogos *online* [Fri12].

³*Fotolog* - é um *blogue* de fotografias que tem mais de 22 milhões de membros em mais de 200 países [Fot12].

⁴*Blogue* - É um *site* que permite a publicação rápida artigos, ou *posts* organizados de forma cronológica inversa, tendo como foco a temática proposta do *blogue* feita pelo seu criador.

⁵*MySpace* - É um *site* de rede social criado em 2003 e vocacionado para a área da musica [MyS12a].

Rússia ou China não têm o mesmo nível de sucesso como podemos ver pela Figura 2.2. Após o *Friendster* e o *MySpace*, o *Facebook*, *Twitter* e o *LinkedIn* nestes últimos anos começaram a ganhar terreno na adesão de novos utilizadores.

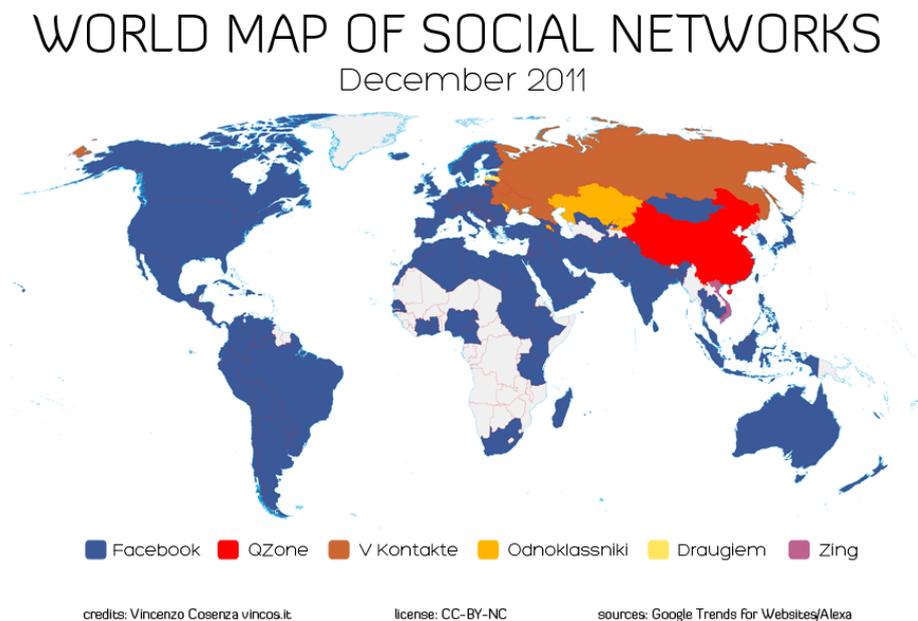


Figura 2.2: Divisão dos participantes pelos *sites* de redes sociais no Mundo em Dezembro de 2011- Retirado de *Google Trends* (Créditos : Vincenzo Cosenza)

2.1.5 Credibilidade da informação nas redes sociais

Os conteúdos criados e consumidos na Internet têm aumentado a um nível impressionante. A título de exemplo, só o *YouTube* [Goo12b] reporta 24 horas de novos vídeos adicionados e mais de 2 biliões de filmes são vistos todos os dias [Min10]. Um grande problema dos *sites* de redes sociais é decidir que actualizações se devem subscrever baseadas na máxima relevância, credibilidade e qualidade da informação recebida. Estes *sites* de redes sociais levam a que os participantes se tornem simultaneamente consumidores e produtores de conteúdos. Enquanto esta mudança na vida quotidiana das pessoas permite aos consumidores de informação mais facilidade em escolherem os conteúdos a seguir, isto torna necessário que os utilizadores descubram, avaliem e seleccionem fontes de informação. Se existe uma informação, idealmente é necessário descobrir a quem essa informação provavelmente poderá interessar. A credibilidade de um utilizador é importante, estando relacionada com a experiência e a confiança [CSP11] que os outros utilizadores têm. A caracterização da experiência de um utilizador pode ser conseguida através da análise das ligações entre os utilizadores ou analisando o conteúdo das mesmas. Existem outras formas de um dado utilizador poder ser classificado como credível através da análise da sua personalidade, por exemplo pela forma como a sua opinião

muda quando vê a opinião dos outros ou quanto um utilizador sabe sobre um determinado tema. A credibilidade pode ainda ser inferida através de um grafo inferindo da própria credibilidade calculado através de indicadores de confiança mútua entre de utilizadores da rede de contactos. Por exemplo, se João confia na Alice e se a Alice confia no Bruno então João confia no Bruno devido a indicação da Alice. Visto isso, podemos dizer que existe um caminho inferido entre o João e o Bruno através da Alice [TD06].

Estes indicadores de credibilidade são pontos de partida de como se poderá chegar a uma solução sobre como calcular a relevância das informações ou mesmo de como calcular a relevância de um utilizador para um determinado grupo. No entanto, a relevância não deverá ser calculada apenas através da credibilidade dos próprios utilizadores mas também de outros factores como forças de ligação criadas através de interacções ou informações explícitas ou implícitas extraídas através dos próprios perfis dos utilizadores.

Com a popularidade ganha pelas redes sociais em todo o mundo, cada vez mais indivíduos se inscrevem nas mesmas criando bastantes perfis de utilizador e produzindo bastante informação. A extracção e análise de informação são dimensões importantes para a determinação da relevância de informação temática de um grupo pelo que apresentarse-á a Secção 2.2 dedicada a essa temática. Também as informações explícitas colocadas pelos utilizadores nos seus perfis são importantes para se perceber que tipos de utilizadores integram nos grupos. Por esse motivo, o conceito de perfil de utilizador, será discutido mais adiante na Subsecção 2.2.1.

2.2 Extracção da informação

Existe um aumento das comunidades virtuais e redes sociais, onde múltiplos utilizadores partilham os interesses e objectivos acedendo aos mesmos espaços virtuais, seja através de interacções implícitas ou explícitas. Este aumento têm motivado o desenvolvimento de ferramentas que suportem e melhorem a utilização destas novas ferramentas. Existem várias formas de detectar em que tipo de conteúdos um utilizador pode estar interessado. Por exemplo no caso de sistemas de recomendação existem diversas propostas para extrair informações relevantes. Muitos sistemas de recomendação são baseados nos perfis de utilizadores, pois contêm informações sobre os seus interesses, o que permite seleccionar informação adequada a cada utilizador, como podemos ver na Subsecção 2.2.1.

2.2.1 Perfis dos utilizadores

Nas várias aplicações que envolvem a interacção com utilizadores, por norma é permitido que cada utilizador crie o seu perfil individual, que o caracteriza univocamente nessas aplicações. Estas aplicações abrangem várias áreas desde as aplicações das redes sociais, ambientes de utilização de sistemas de informação (como nos casos das aplicações interactivas na *Web*), até aos ambientes que, a nível dos sistemas operativos, gerem

as contas de acesso dos utilizadores. Um perfil armazena um conjunto de características que um utilizador associa a uma aplicação, onde coloca a sua informação pessoal. Os dados associados aos perfis podem ser registados de forma explícita, através da divulgação da informação no próprio perfil, ou obtidos de forma implícita, por exemplo através das informações recolhidas nos computadores de cada utilizador [TMB09] ou mesmo de uma forma híbrida [SA09]. Para um utilizador se registar numa nova aplicação, por norma tem de responder a algumas perguntas de forma a serem criadas características para preencher o seu perfil de utilizador. No caso dos *sites* de redes sociais, no seu perfil, um utilizador pode expor informação de diversas formas e associar-lhe certos critérios de visibilidade consoante a privacidade pretendida. A informação presente nos perfis de utilizadores pode ser utilizada para efectuar validações, por exemplo, autenticação e controlo de acesso aos recursos geridos pelo sistema, ou pode, no sentido mais lato, ser utilizada pelo sistema, para configurar ou adequar o conceito da utilização das aplicações consoante as características relevantes do perfil de cada utilizador. Existe um enorme potencial nesta capacidade de um sistema se adaptar ao contexto de utilização de uma aplicação, face à grande diversidade de perfis e/ou face às alterações das suas características ao longo do tempo. Um exemplo típico da concretização de tal potencial encontra-se no caso dos sistemas de recomendação [LVDV99, McC02, MKR04], que se encontram presentes em diversas aplicações e ambientes interactivos acessíveis através da *Web*.

Em relação ao tipo de informação que pode ser disponibilizada nos perfis individuais de cada utilizador, podem existir diferentes conceitos [Pai11] como se descreve seguidamente.

2.2.1.1 Perfil base

Por norma os *sites* de redes sociais têm certas características base para a criação de um perfil tais como o nome, data de nascimento, língua que não sofrem praticamente mudanças ao longo do tempo de vida de um perfil. Apesar de poderem mudar, tal não é comum pelo que ficam definidas como dados do perfil base. Os campos dos perfis que podem ser classificados como base, são geralmente dependentes das aplicações a analisar.

2.2.1.2 Perfil estendido

A ideia de perfil estendido permite lidar com certas informações sobre um utilizador que podem alterar-se mais frequentemente do que as do perfil de base, bem como com informação que só é capturada dinamicamente ao longo do tempo, como resultado das interações do utilizador com a aplicação. Estas informações relacionadas com os interesses mais dinâmicos dos utilizadores, podem ser capturadas através, por exemplo de anotações que ficam associadas ao perfil ou através de informações guardadas sobre as pesquisas realizadas pelos utilizadores. As informações de um utilizador que normalmente mudam com mais frequência estão relacionadas com os seus interesses como por

exemplo, preferências de livros ou programas de televisão favoritos. Também a informação sobre a localização do utilizador, em cada momento, pode ser incluída, como parte do perfil estendido do utilizador. As informações constantes no perfil de estendido são importantes (ou podem ser utilizadas) para guiar os processos de formação de grupos e disseminação de informação.

2.2.2 Filtragem, selecção e extracção de informação

Existem várias formas de seleccionar, filtrar e extrair informação na Internet. Os sistemas de recomendação tentam reduzir o impacto negativo de cada vez existir mais informação nos *sites*. Existem vários modelos para a recomendação de conteúdos, como filtragem baseada em conteúdos (*Content-based filtering*), filtragem demográfica (*Demographic filtering*) e filtragem colaborativa (*Collaborative filtering*).

O modelo de filtragem baseada em conteúdos sugere informação baseada nas opções que o utilizador escolheu no passado. Este modelo, pode sugerir novos produtos para um utilizador adquirir, como acontece com as sugestões da loja virtual da Amazon [Ama12] ou mesmo com lojas virtuais de filmes, que baseado no género de filmes comprados no passado pelo utilizador, sugere novos filmes dentro do mesmo género [SA06].

O modelo de filtragem demográfica (*Demographic filtering*) agrupa utilizadores baseado em algumas semelhanças em atributos pessoais como o sexo, a idade, uma bebida, ou a cidade como acontece com a aplicação *Lifestylefinder* [Kru97].

O nosso objectivo de calcular a relevância da informação, vai-se basear num modelo de filtragem colaborativa, pelo que esta forma de filtragem vai ser descrita em mais detalhe na Subsecção 2.2.3.

2.2.3 Filtragem colaborativa

Uma técnica bastante comum para recomendar algum conteúdo na Internet é através de filtragem colaborativa de conteúdos. Por exemplo é bastante comum existirem *sites* que recomendam, através da opinião de outros utilizadores, jogos como por exemplo o *IGN* [UK12] ou mesmo filmes como *IMDB* [IMD12]. Utilizando dados de diversos utilizadores e comparando-os com um determinado utilizador, é possível sugerir informação mais adequada a esse utilizador através das semelhanças. Existem algoritmos de agrupamento [TSK05] responsáveis por descobrir as semelhanças entre utilizadores como o *K-means* [KMN⁺02], que agrupa consoante o número de agrupamentos (*clusters*) pré definido utilizando a distância média, *Fuzzy c-means* [AYFM99], que agrupa mas permite que um conjunto de dados possa pertencer com um certo grau a vários agrupamentos distintos, ou *DBScan* [Est96], algoritmo baseado em densidade, que encontra agrupamentos usando estimativa da distribuição densidade do nós respectivos. Existem também vários algoritmos que tentam prever, por exemplo o que um grupo de utilizadores irá ou não gostar através da intersecção das listas de interesses com lista de desacordos como Roy em [BRAYC⁺10]. Se um conteúdo pode ter propriedades associadas, pode-se verificar

as semelhanças entre conteúdos, como se faz com os utilizadores. Os conteúdos podem ter anotações (*tags*) associadas, que permite uma categorização dos mesmos. Um sistema que permita anotações a conteúdos, permite que um determinado conteúdo possa ter anotações permitindo identificar o tema desse conteúdo. Esta marcação é benéfica para suportar mecanismos de optimização de pesquisas, classificação de conteúdos e até detecção de conteúdos indesejados, ou seja marcados como lixo electrónico. Através da junção de várias anotações associadas a conteúdos e da junção das propriedades dos perfis de utilizador, se pode descobrir a que tipo de utilizadores um determinado conteúdo pode interessar. Estes ambientes onde pode existir uma marcação de conteúdo colaborativa, permite que os utilizadores marquem os conteúdos de acordo com as suas próprias perspectivas ou sentimentos. No entanto, os utilizadores têm diferentes perspectivas acerca dos conteúdos. Por exemplo, um utilizador pode gostar bastante de uma notícia publicada num determinado grupo, enquanto outros podem discordar totalmente. Baseadas nas anotações que cada utilizador faz de cada conteúdo, podem surgir comunidades para qual os conteúdos interessam. No caso de Xie [XLC12] marcava-se um conteúdo como interessante para uma determinada comunidade, baseado na razão entre a quantidade de anotações que esse conteúdo tinha de um determinado tipo e a quantidade total de anotações. No caso de Almeida [AA04], uma dada pesquisa textual, poderia retornar diferentes resultados dependendo do contexto onde um utilizador estava inserido, utilizando um modelo através da informação da comunidade como fonte para personalizar as pesquisas. Mecanismos de anotações para identificação de conteúdos podem ainda abrir caminho para a área dos grupos baseados em interesses. Por exemplo, utilizadores que tenham feito muitas anotações em comum, têm muitos interesses em comum e podem formar grupos. Elizeu [SNCA⁺09] concluiu através do seu estudo com a ferramenta *CiteULike*⁶ e *Connontea*⁷, que os utilizadores com mais anotações em comum, tinham mais probabilidade de estar inscritos nos mesmos grupos. Este assunto da formação de grupos irá ser abordado mais detalhadamente na Secção 2.3. Muitas vezes a filtragem e identificação dos temas da informação pode não ser fácil, pois pode não existir um sistema de anotações ou mecanismos de identificação como os meta-dados que facilite essa tarefa. Nesses casos muitas vezes são utilizados analisadores de conteúdos. Na Subsecção 2.2.4, abordaremos este tema de forma mais descritiva de forma a percebermos as formas como podemos identificar os conteúdos.

2.2.4 Análise de conteúdos

A análise de conteúdos [Con12b] é uma metodologia usada no estudo no conteúdo de textos e comunicações, normalmente utilizada nas ciências sociais e humanas mas também em diversas áreas científicas, tais como na ciência computacional. Esta metodologia é utilizada muitas vezes para analisar informação, de forma a identificar os tópicos

⁶CiteULike- É um *site* baseado no conceito de marcação de conteúdos através dos utilizadores com o objectivo de promover e partilhar investigações científicas[Cit12].

⁷Connontea - Gestor de referências *online* para clínicos e cientistas [Con12a]

que são partilhados em um ambiente colaborativo. Os métodos utilizados têm sofrido alterações favoráveis não só devido aos progressos linguísticos mas também devido às evoluções computacionais.

Estes métodos permitem a construção de bases de conhecimento analisando a frequência de certos termos, certas construções, relações entre palavras ou mesmo referências num dado texto. Para podermos analisar conteúdos, em aplicações e serviços na Internet é possível utilizar extractores de conteúdos que são ferramentas que automaticamente analisam uma data informação estruturada ou não, para conseguirmos obter informação sobre os utilizadores, não só sobre o comportamento destes mas também sobre o que realmente lhes interessa. Este tipo de aplicações utilizam conteúdos, em que a sua análise tem por base ficheiros, texto, endereços e *HTML*. A escolha de extractores de conteúdos para uma dada aplicação será realizada de forma a que consiga satisfazer todas as necessidades da aplicação. No caso deste trabalho, a escolha dos extractores de conteúdos que poderiam servir para a aplicação a ser desenvolvida, teve em conta se era uma aplicação gratuita ou não e a qual era a sua forma de funcionamento. Referem-se seguidamente os casos de duas ferramentas de análise de conteúdos, que foram consideradas no contexto deste trabalho.

2.2.4.1 *Open Calais*

O *Open Calais* [Ope12a] é um serviço que automaticamente extrai informação semântica de páginas *Web* ou textos que podem ser usados na *Web* semântica [W3C12a]. Com este serviço é possível extrair meta-dados semânticos de conteúdos fornecidos em inglês, identificando as categorias de documentos, anotações sociais e entidades presentes, como se pode verificar na Figura 2.3. Para este efeito, a ferramenta recorre a métodos de processamento de linguagem natural, aprendizagem automática e outros métodos. Embora seja possível apresentar textos em francês e espanhol a sua total funcionalidade só acontece com a língua inglesa.

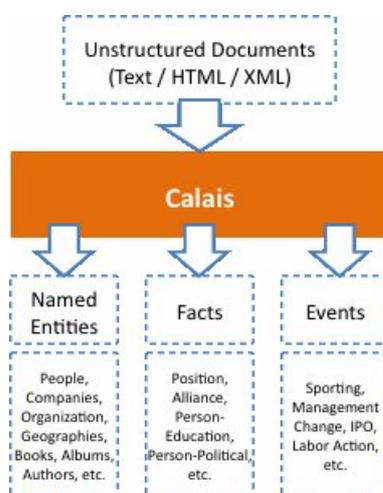


Figura 2.3: OpenCalais

2.2.4.2 *AlchemyAPI*

AlchemyAPI [Alc12b] é um serviço acessível pela Internet, utilizado por programadores de aplicações, que analisa conteúdo não estruturado através da análise de linguagem natural e identifica entidades, locais, factos, línguas, autores, entre outros tópicos. A tecnologia utilizada neste serviço baseia-se no processamento estatístico da linguagem natural, em algoritmos de aprendizagem automática usados para a análise de conteúdos e extracção de meta-dados. Este serviço tem uma utilização livre e gratuita até às 1000 chamadas ao seu *webservice* por dia, sendo, porém, necessário pedir uma chave para a aplicação poder executar essas chamadas. Através da *API REST* [Ric06] ao *webservice* do *AlchemyAPI* [Alc12b] pode-se obter respostas em formato *XML* [XML12], *JSON* [JSOto], *RDF* [RDF12], *REL-TAG Microformat* [REL12] em [*XML-embedded*] e [*raw*]. Para além da interface *REST*, é também possível usar esta ferramenta através das interfaces de programação de linguagens como *Java*, *C/C++*, *C#*, *PHP* [ABC⁺12] ou *Python* [Pyt12] e seu respectivo apoio ao desenvolvimento (*SDKs*). O serviço *AlchemyAPI* consegue detectar várias línguas através de texto, *HTML* ou conteúdo *Web*. Este consegue identificar mais de 95 línguas diferentes [Alc12a] com bons resultados, facilmente categorizando ou filtrando o conteúdo baseado na língua em que é escrito.

2.3 Grupos de utilizadores

Através da análise das redes sociais e dos grupos sociais podemos referir duas formas distintas de adesão a um grupo, sendo estas diferenciadas pela forma como o sistema ou aplicação gerem o “ciclo de vida” de um grupo.

2.3.1 Grupos explícitos

Grupos explícitos referem-se a conjuntos de utilizadores que se inscrevem em grupos por sua própria iniciativa, ou seja, são grupos cujo processo de adesão é controlado explicitamente pelos utilizadores. Em geral, nos grupos explícitos, as acções de saída de um utilizador de um grupo bem como as de eliminação ou criação de um grupo, são também desencadeadas em resposta a um pedido explícito do utilizador.

2.3.2 Grupos implícitos

Grupos implícitos são formados com base em informações derivadas a partir do comportamento que os utilizadores vão exibindo (manifestando) através dos seus hábitos, características, interesses e partilhas [PMC11, MCCC09]. Nestes grupos, os utilizadores não necessitarão de se inscrever para poderem receber informação, pois a sua adesão é inferida através do sistema de gestão de grupos. Por exemplo, se um utilizador está sempre a interagir com a página de desporto e com a página da faculdade então este poderá ter interesse na página de desporto da faculdade. Implicitamente um utilizador poderá ser adicionado a esse grupo de desporto da faculdade se assim o pretender. O sistema

pode inferir automaticamente esses grupos através de diversas técnicas, que iremos referir adiante (na Secção 2.3.3).

No contexto das redes sociais existem variantes do conceito de grupo implícito que consideram que cada grupo é uma combinação de um ou mais contactos sociais com quem o utilizador interage [RBDD⁺10], inferido assim o grafo social. Por exemplo, no caso de Palsma [Pal10], através do estudo das mensagens escritas, enviadas através do *Twitter* consegue-se inferir em que grupos poderia um utilizador ficar implicitamente inscrito.

No trabalho desenvolvido anteriormente por Pais [Pai11], desenvolveu-se uma ferramenta (TIG), para a plataforma *Moodle* [Moo12], em que se consegue implicitamente calcular grupos, através de informação obtida sobre o comportamento de cada utilizador. O objectivo era o de disponibilizar processos e mecanismos genéricos de forma a se poder obter uma ferramenta flexível e útil aos utilizadores em diferentes tipos de aplicações *Web*. Esta ferramenta baseou-se então na análise de informação textual que descreve o comportamento dos utilizadores extraindo informação sobre os seus interesses, características ou interacções. Este processo de extracção era realizado de uma forma dinâmica para se detectarem as mudanças de interesses ou as próprias características. Também as interacções entre utilizadores eram indicadores das suas proximidade e afinidade mútuas, podendo assim ser usados como informação para a criação de grupos implícitos. A análise de conteúdos relacionados com os utilizadores era da responsabilidade de um componente externo, sendo neste caso o *OpenCalais Webservice* [Ope12b], o qual podia ser substituído por outro que desempenhasse funções similares. Existindo um extractor de conteúdos e analisando os grupos e os conteúdos partilhados, podem detectar-se mudanças relacionadas com o tema ou o motivo pelo qual o grupo foi criado. A utilidade de cada grupo pode mudar ao longo do tempo consoante as diferenças na vida social que ocorrem com o utilizador. Por isso qualquer cálculo realizado para a criação de grupos implicitamente deve ser realizado de forma dinâmica. Apesar do sucesso da ferramenta TIG [Pai11], a validação da mesma ferramenta teve alguns problemas devido ao número reduzidos de utilizadores e à limitada actividade de utilização na própria plataforma. Um problema que não foi resolvido com esta ferramenta resultou do facto de criar muitos grupos com um número pouco significativo de utilizadores o que comprometia a eficácia desta ferramenta. A identificação desses problemas reforça o interesse em calcular a relevância dos grupos para os utilizadores e das próprias informações usadas para a sua criação, como forma de evitar aquele tipo de problemas. A limitação daquele trabalho, devida ao reduzido número de utilizadores, necessitava também de ser resolvida, considerando um universo de mais utilizadores, não só para melhor testar e afinar a solução, como também para suportar inquéritos de satisfação, permitindo caracterizar os níveis de satisfação dos utilizadores com a aplicação apresentada.

2.3.3 Critérios de criação dos grupos

Após a distinção entre as várias formas de constituir e gerir grupos vamos identificar os principais critérios para a criação implícita de grupos. O grupos podem ser criados baseados nas interações comuns entre um conjunto de utilizadores como se apresentará na Subsecção 2.3.3.1 e também através de características e interesse como se apresentará na Subsecção 2.3.3.2 .

2.3.3.1 Grupos criados através de interações

O conceito da força das ligações entre utilizadores, *tie-strength*, proposto por Mark Granovetter [Gra73], define a força de uma ligação com base no efeito de combinação da quantidade de tempo despendido, a intensidade emocional e os serviços recíprocos com um utilizador. Esse conceito foi progressivamente aperfeiçoado tendo surgido a caracterização de ligações fortes e ligações fracas. Uma ligação forte está frequentemente associada por exemplo, a utilizadores com quem nos relacionamos como é o caso dos amigos. Ligações fracas têm forças de ligação de fraca intensidade, como a que temos com meros conhecidos com quem não existem laços de afinidade. A caracterização das forças de ligação segundo Gilbert [GK09] baseia-se em sete diferentes características :

- Intensidade - A intensidade é calculada através de variáveis, registando os números de ocorrências de indicadores, tais como palavras partilhadas, comentários feitos sobre fotos ou vídeos, ou mensagens trocadas;
- Intimidade - A intimidade entre utilizadores é calculada através de variáveis como o número de amigos, quantos dias passaram desde a última comunicação ou se existem fotos onde apareçam juntos e qual o estado social do utilizador;
- Duração - A duração é calculada através dos dias desde a primeira comunicação;
- Serviços recíprocos - Os serviços recíprocos são calculados através de variáveis como os *links* partilhados através da *Wall*⁸ e aplicações em comum;
- Suporte emocional - O suporte emocional é calculado através de respostas positivas ou negativas na *Wall* e na caixa de mensagens do utilizador;
- Distância social - Esta componente é calculada através de variáveis como a diferença de idades, número de ocupações diferentes e tipos de política defendidas;
- Estrutura - A estrutura é calculada através de variáveis como o número de amigos em comum, grupos em comum ou interesses em comum.

⁸Uma *Wall* é uma secção no perfil do *Facebook* onde os utilizadores podem partilhar informações. Esta *Wall* é publica segundo as políticas de privacidade, onde os outros utilizadores podem gostar, comentar ou partilhar conteúdos [Wal12].

Através destas características, Gilbert [GK09], calculou a força de cada ligação associada a cada utilizador. Através do cálculo das forças podemos formar grupos, agrupando os utilizadores que estão no mesmo nível ou seja têm o mesmo nível de força de ligação. Embora seja possível calcular as forças das ligações, no caso proposto por Gilbert, tal está limitado ao ambiente de teste por ele utilizado, o *Facebook* pelo que poderá não ser adaptado directamente a outros contextos de utilização.

Roth [RBDD⁺10], através de propriedades de troca de mensagens entre utilizadores e grupos, calcula uma métrica para classificar as ligações entre utilizadores através da frequência, frescura e direcção das interacções entre eles. Através destes cálculos para classificar as ligações entre utilizadores, infere-se uma rede social com as diferentes forças de ligação entre utilizadores. Um utilizador ou grupo com o qual um utilizador mais frequentemente troca mensagens terá uma força maior que outros em que isso não acontece. Através da junção destas três características, frequência, frescura e direcção, pode-se calcular a relevância de cada interacção e assim criar grupos [RBDD⁺10]. Através da criação do grafo social e das suas respectivas forças entre os utilizadores, o algoritmo de sugestão de contactos descrito em [RBDD⁺10] devolve, através de um contacto, um conjunto de contactos que estão ao mesmo nível que a própria *seed*⁹. Este grupo pode ser alterado ao longo do tempo segundo as mudanças de força que cada utilizador tem. Neste caso, por exemplo, um utilizador é identificado pelo seu endereço de e-mail e os nós da sua rede serão outros endereços de e-mail. No caso de se estar a escrever um e-mail e ser colocada uma *seed* a quem queremos enviar o e-mail serão automaticamente indicados quem são os utilizadores que estão ao mesmo nível que a própria *seed*. No caso de escolhermos um utilizador que não estava nesse grupo inferido, o próprio algoritmo identifica e avisa o utilizador de se não terá escolhido um utilizador errado. Estas funcionalidades acontecem no *Google Mail* [Goo11b].

2.3.3.2 Grupos criados através de características e interesses

Os grupos podem ser inferidos através das propriedades dos próprios perfis de cada utilizador como as características dos perfis base ou os interesses dos perfis estendidos das aplicações. Por exemplo, podem inferir-se grupos através das tarefas que estão a desempenhar, através das pesquisas efectuadas [Pai11] ou que tipo de conteúdos estão arquivados no computador [TMB09]. A ferramenta *IGM (Implicit Group Messaging)* [CQL08] entrega mensagens a grupos implícitos que foram criados através de características específicas de utilizadores. Esta ferramenta permite que, sem que um utilizador saiba quem são os utilizadores que irão consumir essa informação, a possa enviar a apenas aqueles utilizadores num ambiente *peer-to-peer*.

Também através de informações retiradas dos perfis dos utilizadores se constroem os sistemas de recomendação [Bur07]. Estes sistemas de recomendação fazem sugestões

⁹Seed - Endereço e-mail inicial para o cálculo do grupo.

aos utilizadores consoante as suas propriedades. Por exemplo, estes podem sugerir o que comprar consoante um produto seleccionado como acontece na *Amazon* [Ama12] ou sugestão de jogos consoante os jogos que foram anteriormente jogados ou que estão em curso como acontece no *Steam* [Ste12].

2.3.4 Perfis de grupos

Através da caracterização de formação de grupos (implícitos ou explícitos) é possível também criar um perfil para o próprio grupo, que o identifica de forma unívoca na própria aplicação. Esses perfis expõem as diferentes propriedades e características de um grupo, como os interesses que levaram à sua formação, quais os utilizadores que dele fazem parte e respectivas interacções e partilhas de informação nesse grupo. O *Facebook*, por exemplo, permite a criação de grupos explícitos, onde é permitido definir um nome para o grupo, definir se o grupo é privado ou público e convidar no momento da criação contactos para aderirem ao grupo [Fac11b]. Após a criação é permitido aos administradores do grupo editarem o mesmo, atribuindo ainda uma fotografia, uma descrição e algumas regras de segurança/privacidade. No caso do *LinkedIn*, ao criarmos um grupo podemos atribuir uma imagem, um nome e que tipo de grupo será (grupo de uma organização, grupo profissional, de uma conferência, entre outros). Para além disto é possível inserir um sumário do que trata o grupo, uma descrição mais detalhada do mesmo, um endereço de um *site*, que tipo de acesso tem o grupo (privado ou público), assim como definir propriedades de idiomas como a linguagem e a localização geográfica [Lin12b].

Todas estas características tornam os grupos sociais numa ferramenta útil e interessante para a disseminação de informação numa rede social de forma organizada, englobando de uma só vez um número possivelmente elevado de utilizadores com interesses comuns.

2.4 Panorama dos conceitos de grupos em *sites* sociais

Actualmente, existem diversos serviços e *sites* de redes sociais, sendo o *Facebook* (Figura 2.2) o mais popular. Para cada um dos sistemas considerados, descrevem-se a seguir as suas principais funcionalidades e de que modo são suportados os conceitos relacionados com a organização dos utilizadores em grupos. Neste estudo também procuramos identificar as funcionalidades a que cada um dos sistemas recorre para avaliação da relevância da informação que processa, bem como em que medida a utiliza no processo de formação, gestão ou disseminação de informação para grupos de utilizadores. De forma a ter uma visão alargada de como funcionam os *sites* de redes sociais, nesta secção vamos descrevê-los de forma genérica para se compreender o seu funcionamento.

2.4.1 *Facebook*

O *Facebook* [Fac11a] é um *site* de rede social fundado em 4 de Fevereiro 2004 por Mark Zuckerberg, Dustin Moskovitz, Eduardo Saverin e Chris Hughes, ex-estudantes da Universidade Harvard, tornando-se numa das redes sociais mais usadas do momento. A sua principal missão é dar às pessoas o poder de partilhar informação e dar a possibilidade de expandir as suas fronteiras físicas fazendo com que o mundo seja mais unido. O *Facebook* permitiu que os seus utilizadores pudessem criar perfis públicos para todos os utilizadores e foi a grande característica que o distinguiu dos outros serviços. Outra ferramenta que o diferenciou, foi o facto de o *Facebook* ter a habilidade de permitir que programadores pudessem criar novas aplicações que depois poderiam ser personalizadas e adicionadas nos perfis dos utilizadores. Em relação aos grupos o *Facebook* apresenta algumas características que anunciamos em seguida.

O *Facebook* permite a criação de vários grupos que podem ser vistos como filtros gerados através de atributos, que estão associados aos perfis dos utilizadores, habitualmente designados por listas de grupos. Por omissão, coloca nessas listas os utilizadores que têm atributos específicos no perfil, sendo, no entanto possível adicionar outros utilizadores. Por exemplo, na lista família por omissão, irão ficar os utilizadores que têm um atributo familiar como mãe, pai ou irmão. Os próprios utilizadores podem criar estas listas e usá-las para receberem e enviarem notificações a pessoas por quem têm interesse. Estas listas permitem a criação, edição e remoção, tanto das listas como dos utilizadores que fazem parte delas.

O *Facebook* permite também a criação de grupos explicitamente pelos próprios utilizadores. Para que um utilizador faça parte destes grupos tem de existir um convite por parte dos administradores do próprio grupo. Dentro destes grupos existem várias funcionalidades como se pode ver na Tabela 2.1.

Funções permitidas nos grupos do *Facebook*

Chat em grupo
 Administrar documentos do grupo
 Administrar álbum de fotos do grupo
 Editar características do grupo
 Envio de mensagens a membros
 Gerir utilizadores do grupo
 Administrar grupo

Tabela 2.1: Operações disponibilizadas pelo grupo no *Facebook*

Em termos de políticas de utilização dos dados o *Facebook* tem algumas normas de utilização [Fac11b]. Respeitar a privacidade dos utilizadores e os próprios utilizadores são duas das mais importantes ali referidas. Todas as aplicações que utilizem dados dos utilizadores necessitam de ter sempre uma autorização para tais dados poderem ser usados. Ao utilizar as *APIs* do *Facebook* temos de ter em conta que interfaces e respectivos campos são necessários para podermos integrar uma aplicação no *Facebook*. Para proceder à

autenticação e autorização dos utilizadores, o *Facebook* utiliza o *OAuth 2.0* [Aut12b]. O *Facebook* tem ferramentas especializadas para o desenvolvimento de aplicações tais como *PHP SDK* para a parte do servidor e *JavaScript SDK* [Fac11c] para a parte do cliente. O *JavaScript SDK* serve para permitir o acesso a todas as ferramentas do *Graph API*¹⁰ e às ferramentas dos *Dialogs*¹¹ via *Java Script*. É necessário usar *XFML*¹² para poder interagir com os *plugins* sociais e com as janelas (*Canvas*) do *Facebook*. O *PHP SDK* [Fac12i] é a ferramenta de ajuda à implementação das funcionalidades do *Graph API* e de autenticação nas aplicações *Web*. O *Graph API* permite obter várias informações como quais os utilizadores que estão relacionados com um perfil de um utilizador e obter informação base deste, como o primeiro e último nome, género, localidade, identificador único na aplicação e língua escolhida. Para além desta informação base é possível obter informação estendida sobre o utilizador, como os seus interesses, estado civil, vídeos ou mesmo lugares onde o utilizador já esteve. A forma como o *Facebook* classifica a relevância da informação é calculada através das interacções. Com autorização dos utilizadores é possível também obter mais informações, tais como que mensagens estão associadas aos perfis dos utilizadores, álbuns e fotografias, vídeos, anotações (*tags*) e *links*. Pode-se obter informação sobre que utilizadores vão a um evento desde que se saiba o nome deste ou mesmo obter estatísticas sobre determinadas aplicações. Em relação a cálculos da relevância de informação a ser recomendada, o *Facebook* baseia-se em todas as interacções sociais executadas pelo utilizador e pela sua rede de contactos e, desta forma, tenta personalizar cada recomendação a cada utilizador específico [Fac12f]. Por exemplo analisa se a informação já foi vista por contactos da sua rede social e estuda a relevância através desses critérios. Em relação à recomendação de pessoas o *Facebook* tem a funcionalidade "*People you may know*" [Fac12g] que, através da informação da rede de contactos, dos contactos em comum, da informação sobre o local de trabalho ou das habilitações académicas, sugere novos utilizadores a adicionar à rede de contactos do utilizador. Por exemplo, na área dos sistemas de recomendação no *Facebook* existe a aplicação *HappyMovie* [SRGDA11] que organiza eventos para idas ao cinema através da análise da informação obtida através de questionários aos utilizadores e informação das forças das interacções de forma a conseguir sugerir os três melhores filmes para o conjunto de utilizadores inscritos em cada evento.

2.4.2 *Twitter*

O *Twitter* [Twi11a] é uma aplicação lançada a 15 de Julho de 2006, que cria uma rede de informação em tempo real pela troca de pequenas mensagens de 140 caracteres, chamadas de *tweets*. Os utilizadores têm acesso a linhas de tempo, de estado e a objectos

¹⁰*Graph API* - É uma *API* simplificada de aceder aos objectos do *Facebook* como os utilizadores, grupos eventos, páginas, etc [com12].

¹¹*Dialogs* - *Dialogs* são janelas que providenciam uma interface simples para mostrar mensagens aos utilizadores, não exigindo especiais permissões.

¹²*XFML*- É um formato *XML* simples para a troca de meta-dados na forma de hierarquias, às vezes chamado de *Taxonomia* [Pet12]

utilizados pelos utilizadores. Suporta-se a partilha de informação sobre os utilizadores, tal como que perfis de *Twitter* está um utilizador a seguir, quem é que o está a seguir ou em que listas está inscrito esse utilizador. De referir que o conceito do *Twitter* é diferente do *Facebook* pois, por exemplo, as mensagens são sempre enviadas para o perfil do utilizador que envia a mensagem e não existe um grafo social organizado da mesma forma que o *Facebook*. Caso um utilizador queira enviar um *tweet* direccionado a alguém basta colocar uma anotação com o nome do destinatário ou escrever na caixa de texto do destinatário. No entanto, esta mensagem irá aparecer no contexto do perfil de quem enviou a mensagem e não no destinatário. Em relação aos grupos podemos ver que operações são permitidas através da Tabela 2.2.

Funções permitidas no <i>Twitter</i>
Gestão de listas
Estrutura <i>publish/subscribe</i> ¹³
Administração de subscritores ¹⁴
Pesquisa por palavra chave representada por #maispalavra

Tabela 2.2: Operações disponibilizadas pelo *Twitter*

Através de uma *API REST* [Ric06], é possível aceder às primitivas base do *Twitter* [Twi11b]. Utiliza o *Auth* [OAu12] para fazer a autenticação dos utilizadores e usa nas suas *APIs* o protocolo HTTP [FGM⁺99], linguagem XML [XML12], JSON [JSOto], RSS [RSS12] e *Atom* [Ato12] para representar a informação. De forma a recomendar certa informação o *Twitter* usa o conceito de *Streaming* que filtra a informação através de vários atributos, tais como, o número do utilizador, palavra chave ou área geográfica. No caso da sugestão ser sobre que utilizadores seguir, *Twitter* sugere novos utilizadores com base nos seguidores dos outros contactos. Por exemplo, se o Manuel segue a Alice e a Alice segue o João, então o *Twitter* vai sugerir que o Manuel adicione o João.

2.4.3 *LinkedIn*

É uma rede social vocacionada para a área profissional, lançada a 5 de Maio de 2003 [Lin12a]. Nesta ferramenta os utilizadores podem expor o seu *Curriculum Vitae* e podem se candidatar a propostas de emprego colocadas pelas empresas. As empresas por seu lado colocam propostas de emprego e analisam utilizadores que possam corresponder às expectativas da sua empresa. Existe informação sobre as estatísticas de utilizadores, como quem é que vê a rede de um certo utilizador e a que zona esse utilizador pertence. Também mantém números sobre o quão a rede cresce e quem pertence à nossa rede em primeiro, segundo e terceiro grau. Existe um gestor de sectores empresarial interessado

¹³*Publish/subscribe*- É um padrão de envio de mensagens, onde o publicador (publisher), envia mensagens para um determinado grupo de utilizadores que tiveram interesse em receber mensagens (*subscribers*)[PM10].

¹⁴Administração de subscritores - O *Twitter* permite que haja uma gestão dos utilizadores que recebem a minha informação.

no perfil de certo utilizador. O *LinkedIn* permite que sejam utilizados dados dos utilizadores desde que estes autorizem o seu uso. Qualquer dado do *LinkedIn* não poderá ser guardado, a não ser o da identificação do utilizador e qualquer informação que esteja explícita no perfil do utilizador, desde que haja uma componente interface que pergunte ao utilizador se este quer disponibilizar os seus dados.

Existem algumas restrições no acesso aos dados promovido pelo *LinkedIn*. Tudo o que não seja do perfil do utilizador não poderá ser gravado. Em relação a grupos o *LinkedIn* tem as seguintes funções como podemos ver na Tabela 2.3.

Funções dos grupos no <i>LinkedIn</i>
Gestão de grupos
Sugestão de grupos - Baseado nos grupos da rede de contactos
Gestão de grupos em destaque

Tabela 2.3: Operações nos grupos disponibilizadas pelo *LinkedIn*

Esta aplicação está vocacionada para o mundo empresarial e funciona também como um centro de empregos, o que a distingue das outras aplicações. Usa duas *APIs* distintas, uma para a parte do cliente e outra para a parte do servidor. Para a parte de implementação do cliente usa *JavaScript* e para a parte do servidor usa uma arquitectura *REST* através de linguagem *XML* e *JSON* [JSOto]. Usa tecnologia *OAuth 2.0* [Aut12a] para fazer autenticação dos utilizadores do *LinkedIn* e permitir o acesso à informação através de objectos e interacções com os *Plugins*. As *APIs* do *JavaScript* suportam várias anotações que ajudam o *LinkedIn* a extrair informação [Lin11]. Em relação às informações sugeridas ao utilizador, como os grupos de interesse ou novos utilizadores, o *LinkedIn* baseia-se em graus de interacção e em características em comum, como as habilitações académicas ou o local de trabalho [Lin12c].

2.4.4 *YouTube*

O *YouTube* [Goo12b] é uma plataforma que permite ao utilizador publicar e partilhar vídeos em formato digital na Internet. Exibe toda a informação sobre os vídeos partilhados e sobre o utilizadores que a partilharam. Em relação às funcionalidades permitidas pelo *YouTube*, podemos observar a Tabela 2.4.

Operações no <i>Youtube</i>
Não tem nenhum sistema de gestão de grupos
Estrutura <i>publish/subscribe</i>
Popularidade de vídeos
Estatísticas de vídeos
Sugestão de vídeos - Baseado na visualização de outros utilizadores

Tabela 2.4: Operações disponibilizadas pelo *Youtube*

A obtenção de informação através da *API* do *YouTube* passa pelo uso da *DataAPI*. Esta permite aos utilizadores colocarem o reproduzidor de multimédia do *YouTube* nos seus *sites*.

A informação colocada é feita através dos *uploads* de vídeos pelos utilizadores. Tem-se acesso assim a toda a informação sobre os filmes que o *You Tube* contenha e pode personalizar a sua página consoante a informação que tem. As *APIs* de dados da *Google* são baseadas em *Atom* 1.0 [Ato12] e *RSS* 2.0 [RSS12] e usam *Atom Publish Protocol* [Sne06]. Se usarmos a *API* 2.1 podemos usar *Google accounts* para aceder a grande parte das funcionalidades do *You Tube* sendo esta a *API* mais recente. É necessário utilizar uma chave da *Google* para podermos usar as *APIs*. Utiliza o *OAuth* [OAU12] para a autenticação em aplicações *Web* e para o controlo e partilha da informação usa *XML* [XML12] e *HTTP* [You11]. Para calcular toda a informação sobre os vídeos e saber quais sugerir, o *YouTube* não tenta apenas basear-se nos mais populares ou nos últimos a serem inseridos. Para cada filme, consoante os critérios de relevância utilizados como por exemplo, quantas recomendações existiram ou quantas vezes foi visto através de outro filme, é criado um grafo implicitamente que é uma lista de filmes a visualizar através de um filme. Ao contrário de todos os *sites* de redes sociais até agora, o *YouTube* tem um grafo centrado no conteúdo, ou seja num filme em vez de centrado num perfil de utilizador [BSS⁺08].

2.4.5 *Google+*

Google+ é uma rede social criada pela empresa *Google* e lançada em Junho de 2011 [Goo]. A principal base do *Google+* é simplificar o processo de partilha e relacionamento entre pessoas na *Web*, seja o mais semelhante ao da vida real, ou seja, se eu tenho conhecimento de alguém esse alguém também deve ter conhecimento de mim. Em termos de grupos o *Google+* usa o conceito de círculos [Goo11a] que acaba por ser um filtro de informação. A partilha de conteúdos diferentes é feita com pessoas diferentes e os círculos ajudam na divisão da partilha da informação. Os círculos ajudam a organizar todas as pessoas de acordo com as suas ligações sociais da vida real – por exemplo: "família"; "amigos do trabalho"; "amigos de escola"; "ex-alunos". Nesta aplicação é possível personalizar os dados pessoais do perfil para os seus círculos. O círculo "amigos" pode ver os dados de contacto, a localização e informações sobre a sua relação, ao passo que o historial de empregos e formação académica estariam visíveis para o círculo "antigos alunos".

O *Google+* também sugere novas pessoas para serem adicionadas à sua rede social baseado nos critérios da Tabela 2.5 .

Critérios para a sugestão de novos utilizadores no <i>Google+</i>
Ligações em comum com pessoas noutros produtos da <i>Google</i> , como o <i>Gmail</i>
Interacção que tenha tido com outras pessoas noutros produtos da <i>Google</i>
Links que tenha adicionado ao seu perfil
Contas associadas que tenha associado à sua Conta <i>Google</i>
Pessoas nos seus círculos alargados

Tabela 2.5: Critérios para a sugestão de novos utilizadores no *Google+*

Este sistema usa *HTTP* [FGM⁺99] e *JSON* [JSOto] para estruturar a informação. Por

exemplo um utilizador é visto como um objecto descrito em *JSON*. O *Google+* disponibiliza *APIs* que dão toda a informação do perfil de um utilizador. O *Google+* permite assim dessa forma procurar por alguém específico ou por utilizadores que tenham o seu perfil público. Permite detalhar também as actividades e os comentários que um utilizador está a desenvolver. Tal como no *YouTube* é necessária uma chave *Google* para se poder utilizar as *APIs*. É possível obter informação base e estendida sobre os utilizadores. Sobre a sua própria rede não é possível obter nada. Apenas é possível obter que utilizadores estão associados a uma determinada actividade *post*.

2.4.6 *Flickr*

O *Flickr* [Fli12a] é um *site* de hospedagem e partilha de informação como fotografias, ilustrações e vídeos, que permite novas formas de organizar as suas fotos ou vídeos, também caracterizado como um *site* de rede social. O *Flickr* começou em 2004 no Canadá e foi transferido para os USA após a sua aquisição pela *Yahoo* em 2005. O *Flickr* organiza e classifica as fotos com *meta-dados* como informações de licenças, localização geográfica e anotações que são atribuídas às fotografias pelos próprios utilizadores. Permite a criação e partilha de álbuns por todo o mundo, o que incute a troca de informação entre utilizadores. Em termos de grupos o *Flickr* utiliza a sua própria ferramenta de grupos, onde se tem as opções da Tabela 2.6.

Operações sobre os grupos disponibilizadas pelo <i>Flickr</i>
Criar novos grupos
Listar e procurar grupos
Convidar membros para os grupos
Partilhar conteúdos para os grupos
<i>Wall</i> de grupos

Tabela 2.6: Operações sobre os grupos disponibilizadas pelo *Flickr*

Em relação à *API* [Fli12b] do *Flickr*, é necessário obter uma chave para a aplicação que requisita os dados através do *Flickr*. É possível fazer chamadas à *API* do *Flickr* através de *REST*, *XML-RPC* [Ric06] e *SOAP* [SOA12], e obter vários tipos de resposta da mesma forma e ainda em *JSON* [JSOto] e *PHP* [ABC⁺12]. Para fazer a autenticação dos utilizadores, o *Flickr* usa o padrão de mercado o *OAuth 1.0* [OAu12]. Permite a obtenção dos grupos que um dado utilizador está inscrito assim como a informação associada a cada um.

2.4.7 Discussão e comparação entre os *sites* de redes sociais

Nesta secção são apresentadas comparações com base na informação que foi possível retirar a partir das descrições disponíveis das *APIs* das redes sociais. É um facto que existem restrições que dificultam a obtenção de informação sobre as estruturas internas e a implementação das redes sociais devido as diferentes políticas de privacidade. Para suportar esta comparação, apresentam-se várias tabelas com os vários *sites* de redes sociais.

<i>Sites</i> de redes sociais	Nome	Identificador único e língua	Nome do utilizador	Aniversário e género
<i>Facebook</i>	✓	✓	✓	✓
<i>Google+</i>	✓	✓	✓	✓
<i>Twitter</i>	✓	✓	✓	Não tem género
<i>LinkedIn</i>	✓	✓	✓	✓
<i>Youtube</i>	✓	✓	✓	✓
<i>Flickr</i>	✓	✓	✓	Não tem aniversário

Tabela 2.7: Operações disponibilizadas pelas *APIs* a nível da informação base

Podemos verificar que em termos de informações do perfil base todos têm informação que se pode extrair da *API* de cada *site* como podemos ver na Tabela 2.7. No entanto, o *Twitter* não tem o campo do género do utilizador. Em termos de informação do perfil estendido já existem bastantes diferenças entre os vários *sites* de redes sociais. Muitas aplicações não permitem a visualização das mensagens recebidas ou enviadas devido às suas políticas de protecção dos dados (Tabelas 2.8 e 2.9).

<i>Sites</i> de redes sociais	Informação temporal e localização	Interesses e biografia	Mensagens enviadas	Mensagens recebidas
<i>Facebook</i>	Data de criação não pública	✓	✓	✓
<i>Google+</i>	✓	Não tem interesses	X ¹⁵	X
<i>Twitter</i>	✓	Não tem interesses	✓	X
<i>LinkedIn</i>	✓	✓	X	X
<i>Youtube</i>	✓	✓	X	✓
<i>Flickr</i>	✓	✓	X	X

Tabela 2.8: Operações disponibilizadas pelas *APIs* a nível da informação estendida - (I)

Em relação a grupos e à própria rede social que é criada nestas aplicações pelos utilizadores, existem conceitos diferentes para as diferentes redes sociais. Enquanto que o *Twitter* e o *YouTube* têm o conceito de produtor (que publica) e consumidor, as outras aplicações têm o conceito de amigos. No entanto é possível criar uma rede social baseada em ambas as informações disponibilizadas. Em relação às *APIs* só o *Twitter*, *YouTube* e *Facebook* disponibilizam através das *APIs* a sua rede social (Tabela 2.10). No caso do *Google+* apenas é possível ver quem são os utilizadores de uma determinada actividade.

O *Flickr* também poderia ser uma boa base experimental para este trabalho, mas como poderia existir uma necessidade de verificar as interações entre utilizadores, não foi tornado como opção.

¹⁵X - Este símbolo representa que aquele *site* de rede social não suporta a operação indicada.

Sites de redes sociais	Conteúdo das mensagens	Educação	Trabalho ou Organizações	Estado da relação
<i>Facebook</i>	✓	✓	✓	✓
<i>Google+</i>	✓	✓	✓	✓
<i>Twitter</i>	Ultima	X	X	X
<i>LinkedIn</i>	✓	✓	✓	X
<i>Youtube</i>	X	✓	✓	✓
<i>Flickr</i>	Só as publicas	X	✓	✓

Tabela 2.9: Operações disponibilizadas pelas APIs a nível da informação estendida - (II)

Sites de redes sociais	Grupos	Grupo por omissão	Aprovações a conteúdos	Conexões	Filtros
<i>Facebook</i>	✓	✓	✓	✓	✓
<i>Google+</i>	✓	X	✓	Consoante a Actividade	✓
<i>Twitter</i>	X	X	✓	✓	✓
<i>LinkedIn</i>	✓	X	✓	X	✓
<i>Youtube</i>	X	X	✓	✓	✓
<i>Flickr</i>	✓	X	✓	✓	✓

Tabela 2.10: Operações disponibilizadas pelas APIs a nível da informação estendida - (III)

2.5 Resumo

Ao longo deste capítulo foram apresentados vários conceitos e trabalhos já realizados nas diferentes áreas de forma a dar suporte à concepção da ferramenta a ser desenvolvida. Após analisadas as diversas APIs, decidiu-se que devido às várias restrições impostas pelas mesmas APIs de cada site de rede social, o estudo seria focado no *Facebook*. Esta escolha incide no *Facebook* devido à sua facilidade em extrair dados através da sua API e devido ao facto de utilizar um grafo social para classificar as suas interações com base em contactos. Embora outros sites de redes sociais também se baseiem num grafo social, nesses casos o acesso é bastante restrito como no caso do *LinkedIn*, ou o grafo não é formado em relação aos utilizadores mas ao conteúdo como é o caso do *YouTube* que cria grafos a partir de vídeos. O facto de existir um grafo social permite que sejam avaliadas as forças de ligação entre utilizadores e que seja estudada a própria rede social de um utilizador e dessa forma este considera-se um factor bastante importante. De facto *Flickr* também poderia ser uma opção mas visto ser muito específico para fotografias e vídeos e já existirem sistemas de estatísticas para os conteúdos, não achamos que fosse haver um grande impacto em criar através desta plataforma a aplicação. Outro motivo é o facto de a actividade nos grupos no *Flickr* ser um pouco mais reduzida que no *Facebook*, por exemplo. Também o *Twitter* não se tornou opção devido ao seu grafo ser baseado apenas em seguidores e seguidos. Por fim também a juventude do *Google+* quanto ao seu desenvolvimento, tornou a sua escolha difícil. Apesar de ter as funcionalidades de grupos como o *Facebook* ou mesmo *LinkedIn*, a sua API não disponibiliza qualquer forma de extrair informação sobre os mesmos mas apenas sobre os utilizadores.

Como extractor de conteúdos decidiu-se usar o *AlchemyAPI* [Alc12b], pois esta ferramenta consegue reconhecer o português e fazer a extracção de tópicos também nesta língua o que irá facilitar bastante depois nos testes, na usabilidade da aplicação no *Facebook* e na própria correspondência entre os tópicos dos *posts* e dos grupos. O facto deste extractor de conteúdos poder analisar várias línguas diferentes, permite que esta aplicação seja internacionalizada e usada com outras línguas que não apenas o português já que a correspondência é feita na língua usada. No capítulo seguinte abordam-se algumas formas de calcular a relevância da informação na gestão de grupos de utilizadores.



Modelo para a avaliação da relevância em grupos

Neste capítulo desenvolvem-se os conceitos e apresentam-se as ideias que fundamentam a abordagem seguida, para a criação de uma ferramenta que consiga analisar a relevância ou popularidade da informação, suportando a disseminação da informação e a avaliação da relevância da informação e de utilizadores em grupos.

Irá-se-á descrever a ideia geral, quais os componentes que fazem parte do modelo e também quais as informações que irão ser extraídas dos vários tipos de perfis, seja de grupo ou de utilizador. Por fim, apresentar-se-ão os componentes relativos à análise de relevância de conteúdos, que serão discutidos de forma a se perceber que indicadores de interesse serão utilizados e de que forma serão utilizados para o cálculo da importância dos conteúdos.

3.1 Abordagem

Nesta secção apresenta-se um modelo que permite avaliar de forma quantitativa a relevância/popularidade dos conteúdos no contexto dos grupos. Pretende-se que este modelo seja o mais neutro possível, de forma a ser mais heterogéneo em relação à dimensão da semântica para poder ser usado independentemente da plataforma de base onde é usado. Por outras palavras, existindo um conjunto de utilizadores e os seus vários grupos, quer fossem criados de forma implícita ou explícita, pretende-se avaliar os conteúdos que são partilhados no contexto dos grupos de forma a se poder quantificar a relevância das informações partilhadas nos grupos, ao longo de um determinado intervalo de tempo. O intervalo de tempo em que a análise da relevância se processa, pode

ser regulado. No entanto, não se pretende que sejam analisados dados cujo o indicador temporal seja demasiado elevado.

Na Figura 3.1 apresenta-se de uma forma esquemática, os principais componentes do modelo proposto.



Figura 3.1: Abordagem geral do modelo

Uma partilha de conteúdos pode ter vários componentes como uma mensagem, um vídeo, uma imagem ou um endereço *URL*. Relativamente a esses conteúdos partilhados existe associado um leque de acções possíveis que representam a opinião dos utilizadores sobre o conteúdo partilhado, como é o comentário, a partilha para outros utilizadores e o marcador "gosto". Existe um interesse em quantificar o quão os utilizadores de um determinado grupo têm interesse no que os utilizadores partilham, e esse é um grande propósito deste modelo. A este valor quantitativo caracterizamos a relevância de que um utilizador tem para o grupo através da popularidade dos conteúdos partilhados. Neste caso consideramos que a relevância de um conteúdo, será a popularidade que esse conteúdo tem para um grupo. Através desse valor pode-se verificar (ou quantificar) a popularidade de um utilizador no contexto de um grupo. Este valor deve reflectir o grau de satisfação dos membros de um grupo para com a informação partilhada neste como em [BRAYC⁺10]. Quanto maior for o número de membros de um determinado grupo a demonstrarem interesse, através das acções permitidas sobre um conteúdo, maior será o valor da popularidade de um conteúdo. Associados às próprias acções que se podem efectuar sobre um determinado conteúdo, bem como o tempo em que um utilizador não participa num grupo, podemos perceber a actividade de um utilizador no seu grupo e verificar a importância do grupo para o utilizador.

Nas próximas secções vão ser apresentados os elementos que vão ser avaliados e utilizados para a realização deste modelo. Na próxima secção (Secção 3.2) apresentaremos algumas propriedades dos utilizadores e do grupo relativamente aos perfis de utilizadores.

3.2 Perfis de utilizador e de grupo

Os utilizadores e os grupos contêm informações importantes para se identificar as propriedades específicas, por exemplo no caso de um grupo, a sua descrição e para que fim foi esse grupo criado ou, no caso de um utilizador, identificando os seus interesses. Essas informações encontram-se através dos perfis de grupo e dos perfis de utilizador normalmente encontrados em *sites* de redes sociais. Nos dois tipos de perfil pode-se obter informação de cariz diferente consoante seja um utilizador ou seja um grupo de utilizadores e dessa forma há necessidade de separar a explicação dos dois perfis.

3.2.1 Perfil de utilizador

No perfil do utilizador, é possível extrair informações não só através do perfil base mas também através dos perfis estendidos de forma a podermos obter informações relevantes para caracterização do utilizador perante um grupo. Através do perfil estendido do utilizador é possível extrair informação sobre as interações de um utilizador ou seja quais as interações que são feitas e quais as suas características tais como frequência, a frescura e quantidade. O perfil de utilizador também permite obter quais os principais interesses de um utilizador e perceber o seu interesse sobre um determinado assunto, analisando o conteúdo dos documentos ou tópicos que estão marcados no seu perfil. De uma forma explícita, um utilizador pode expor os seus interesses no perfil estendido e nesse caso os campos estão completamente instanciados. Estas características são importantes para a classificação do utilizador em relação a que informação tem ou não mais relevância consoante um determinado tópico. De forma a sistematizar o perfil do utilizador ilustramos a Figura 3.2.

Outra informação importante a extrair do perfil do utilizador é a informação que é pública, extraída através das mensagens públicas. Essas mensagens mostram que informação é que um utilizador quer partilhar com os outros utilizadores, como uma música ou uma fotografia pessoal. Este tipo de informação encontra-se muitas vezes partilhada não só nas páginas pessoais dos utilizadores como também nos grupos em que um utilizador se insere. Desta forma, surge a ideia da partilha de informação através da análise da informação, de forma que, consoante o grupo, se sugira que informação será interessante partilhar ou não. O estudo do perfil de utilizador, no contexto deste modelo não é significativamente utilizado, apesar de em trabalho futuro se poder utilizar estas informações para que a partilha de informação possa ainda ser mais adaptável ao contexto do utilizador.

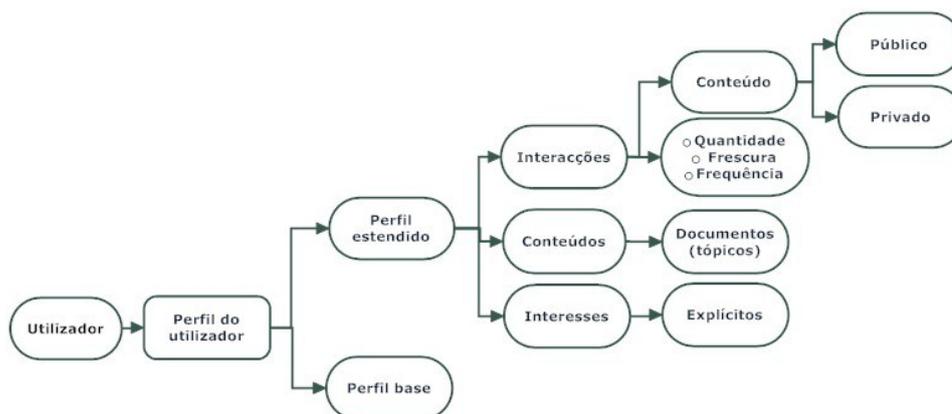


Figura 3.2: Diagrama do perfil de utilizador

3.2.2 Perfil de grupo

Um grupo também tem as suas propriedades, tal como um utilizador, pelo que o conceito de perfil, mostra-se importante neste contexto, no qual um grupo concentra as suas principais características. O conceito do perfil de grupo captura propriedades diferentes das do perfil de utilizador mas que também podem ser extraídas e analisadas. Existem duas dimensões diferentes de informação que podem ser retiradas do perfil de um grupo, conforme se relacionam com o perfil base ou com o perfil estendido.

3.2.2.1 Perfil base

O conceito de informação base no perfil de grupo, define as características ou atributos relacionados com informação sobre a criação do grupo, ou seja como é que o grupo foi criado, quando é que o foi e quem faz parte dele. Neste conceito são considerados dados sobre a rede social correspondente a esse grupo indicando quem são os utilizadores que fazem parte deste. Também no perfil base de um grupo existe o conceito de interesse explícito resultante, por exemplo, da forma como o criador de um grupo especificou explicitamente os interesses associados ao grupo que criou. No caso do grupo ser criado de forma implícita estes atributos são calculados através das características dos utilizadores do grupo [PMC11, MCCC09].

3.2.2.2 Perfil estendido

Dependendo das interacções ou partilha de informação entre os membros, cada grupo apresenta um conjunto de características ou atributos que evoluem ao longo do tempo. Por exemplo, num grupo é possível agregar documentos ou outro tipo de informação partilhada, que fica associada aos conteúdos acessíveis aos membros do grupo. Outra característica do grupo que pode evoluir dinamicamente, são as interacções por mensagens entre membros, por exemplo através de *posts*, sendo que os seus atributos podem ser medidos por indicadores quantitativos, como por exemplo quantidade, frescura ou

frequência das interações. A informação que descreve estas características dinâmicas do grupo pode ser utilizada para ajudar a avaliar a relevância da informação partilhada no contexto de grupo.

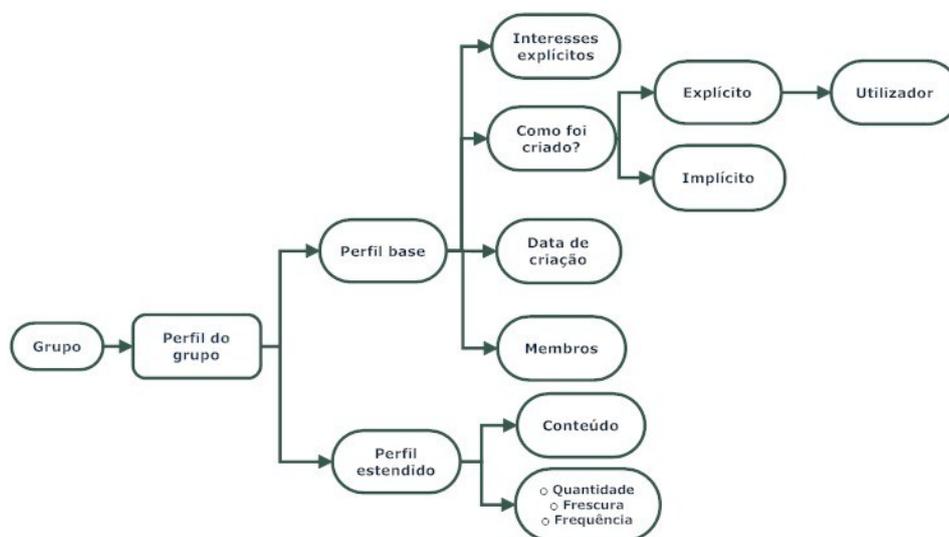


Figura 3.3: Perfil grupo

De modo a sistematizar a forma como as informações sobre um grupo podem ser organizadas podemos ver a estrutura do perfil de grupo através da Figura 3.3. No entanto será necessário ver que informações são possíveis de obter utilizando a rede social escolhida para o efeito. Pelo estudo realizado no capítulo anterior verificamos que a maior parte das redes sociais tem por base um conjunto comum de características para os grupos.

Após esta aproximação sobre os perfis de utilizador e do grupo, é importante identificar os métodos que permitam extrair informação que caracterize a relevância ou popularidade dos conteúdos partilhados nos grupos. Na Secção 3.3 explica-se detalhadamente todo esse processo de análise, quais os procedimentos utilizados e os componentes envolvidos. Devido à interacção existente entre perfis de grupo e perfis de utilizador, torna-se interessante saber que informação tem tido bastante popularidade e quem foram os utilizadores que partilharam esses conteúdos.

3.3 Popularidade dos conteúdos

Com base na forma como os utilizadores podem expressar as suas opiniões sobre os conteúdos é importante definir uma forma de quantificar a preferência dos utilizadores sobre uma determinada informação partilhada. Os indicadores de interesse por um conteúdo da parte de um utilizador são importantes para se fazer a quantificação do interesse que houve por um conteúdo partilhado. Nesta secção explica-se quais os indicadores de conteúdos que serão utilizados (Subsecção 3.3.1) e como será calculada a popularidade de um determinado conteúdo partilhado (Subsecção 3.3.2).

3.3.1 Indicadores de interesse em redes sociais

Grande parte das redes sociais produzem informação de forma idêntica, seja através de mensagens, comentários, partilha de fotos ou vídeos, ou mesmo através da marcação do interesse em algum conteúdo específico, por exemplo, colocando-o como favorito.

Associados à informação partilhada por um utilizador, existem normalmente indicadores de interesse, que podem ser identificados se um utilizador tiver interesse pelo conteúdo, seja de uma forma positiva ou de uma forma negativa. Por exemplo, uma fotografia artística partilhada através do *Flickr*, tanto pode ter um comentário positivo a dizer que a imagem é "magnífica", como pode ter um comentário a criticar a fotografia, por exemplo, por ser demasiado amadora para ser publicada. Nos dois casos foi revelado um interesse por parte dos utilizadores que comentaram a fotografia. Os indicadores de interesse que se consideram, são os seguintes:

- Comentário - O comentário permite que exista uma pequena mensagem a criticar o conteúdo;
- Partilha de um endereço *URL* - Esta partilha permite que seja considerada a quantidade de vezes que um utilizador partilhou aquele conteúdo, assumindo que o partilhou pois o considerou interessante;
- Gosto - O gosto permite que conteúdos sejam marcados com a indicação que um utilizador gostou desse conteúdo.

Através destes elementos será possível quantificar a opinião dos utilizadores sobre um determinado conteúdo partilhado consoante o número de indicadores encontrados. Estes indicadores de interesse podem ser encontrados não só nos perfis de utilizadores mas também nos perfis dos próprios grupos. Neste caso serão avaliadas as informações relativas à partilha nos grupos.

Para cada conteúdo partilhado num determinado grupo G , são recolhidos indicadores quantitativos de interesse, segundo os diversos tipos considerados, ou seja C (comentário), S (partilha/*share*) ou L (gosto/*like*) produzidos por todos os utilizadores pertencentes a esse grupo para proceder ao cálculo da relevância de um determinado conteúdo no ambiente do grupo.

Estes parâmetros correspondem à Tabela 3.1.

Indicadores de interesse - Variáveis
N_c : Soma do número de comentários feitos por todos os utilizadores num grupo em relação a um determinado conteúdo partilhado.
N_s : Soma do número de partilhas feitas por todos os utilizadores num grupo em relação a um determinado conteúdo partilhado.
N_l : Soma do número de gostos(<i>likes</i>) feitos por todos os utilizadores num grupo em relação a um determinado conteúdo partilhado.

Tabela 3.1: Indicadores de interesse - Variáveis

Para todos os conteúdos partilhados em cada grupo, é importante fazer o cálculo dos indicadores de interesse e analisar os seus valores dentro desse grupo, para que se

possam identificar os comportamentos característicos comuns a esse grupo. A média por indicador de interesse é calculada por forma a avaliar-se se um conteúdo está com uma popularidade acima ou abaixo do que se esperava, ou seja, do que é usual nesse grupo. Isso é calculado através da soma de todos os indicadores de interesse de um determinado tipo (Comentário (C), Partilha (S) e Gosto (L)), pelo número de conteúdos partilhados. Através da Tabela 3.2, podemos representar as variáveis responsáveis pela média dos indicadores de interesse.

Média dos indicadores de interesse nos grupos
A_c : Média de comentários realizados num grupo por um conteúdo partilhado.
A_s : Média de partilhas realizadas num grupo por um conteúdo partilhado.
A_l : Média de gostos realizados num grupo por um conteúdo partilhado.

Tabela 3.2: Média dos indicadores de interesse nos grupos

Ao calcular, para cada grupo, os seus próprios valores, expõem-se características do comportamento do grupo, baseadas na actividade dos utilizadores, permitindo uma adaptação ao tipo de cada grupo, como se discutirá mais adiante.

A utilização dos valores médios para aqueles indicadores permite estabelecer valores de referência do comportamento comum de cada grupo e assim identificar/expor/revelar os conteúdos que evidenciam maiores índices de popularidade. Juan Cameron[CLT11], quando calculava o peso¹ de grupos de utilizadores numa base de dados, fazia uma mesma analogia. Tentava calcular a força de ligação de toda a base de dados e depois calculava o rácio de um grupo específico utilizando o peso do grupo e o peso total para identificar que grupos tinham uma força de ligação maior. Existe um rácio que relaciona a popularidade que um conteúdo normalmente tem no grupo com a popularidade do conteúdo que está a ser calculado, para permitir evidenciar o valor de algo que está acima da média.

3.3.2 Relevância dos conteúdos

As expressões seguintes podem ser usadas com os indicadores de interesse descritos na Secção 3.3.1, o que irá permitir classificar a relevância de conteúdos partilhados num determinado grupo. Estes indicadores de interesse são utilizados pois estão relacionados com os comportamentos dos utilizadores em ambientes onde os conteúdos são partilhados (*sites* de redes sociais) e nas formas disponibilizadas para os utilizadores interagirem e partilharem informação. Assim sendo, a relevância de um conteúdo partilhado, pode ser dada pela Expressão 3.1:

$$R_c = \left(\frac{N_c}{A_c} * W_c\right) + \left(\frac{N_s}{A_s} * W_s\right) + \left(\frac{N_l}{A_l} * W_l\right) \quad (3.1)$$

De notar que os símbolos N_c , A_c , N_s , A_s , N_l e A_l , designam os indicadores definidos anteriormente nas Tabelas 3.1 e 3.2.

¹Peso de um grupo neste contexto refere-se à força de ligação entre membros de um determinado grupo.

Para o cálculo da relevância de um conteúdo existem 3 diferentes constantes que representam o peso de cada componente que um conteúdo pode receber.

1. W_c : Peso dos comentários na fórmula do cálculo da relevância;
2. W_l : Peso dos gostos na fórmula do cálculo da relevância;
3. W_s : Peso das partilhas na fórmula do cálculo da relevância.

Sendo que $W_c + W_l + W_s = 1$

O somatório do peso dos indicadores de interesse será no máximo 1 para que quando multiplicado pelo valor de rácio, no máximo o seu valor seja 1.

A regulação dos valores destes pesos é um aspecto que ficou em aberto neste trabalho, sendo de notar a possibilidade de vir a permitir que a contribuição de cada componente seja regulável, por exemplo, consoante os padrões de actividade e os contextos de utilização dos grupos. Como associada a um conteúdo partilhado existe informação sobre quem partilhou essa informação, podemos verificar não só que relevância é que um conteúdo tem, mas também quais são os utilizadores que partilham informação mais relevante para o grupo. Dessa forma torna-se interessante calcular a relevância dos utilizadores para os grupos e perceber quem é que efectivamente partilha informação importante para o grupo. Na Secção 3.4 explica-se a forma como se calcula a importância dos utilizadores para os grupos.

3.4 Relevância de utilizadores para os grupos

A actividade nos grupos em redes sociais pode ser bastante frequente ou intensa, porém, a simples análise da frequência de partilha de informação no grupo, não permitirá concluir que a informação partilhada é relevante. Dessa forma e baseado na relevância dos conteúdos partilhados, pode-se identificar quem efectivamente são os utilizadores relevantes num contexto de um determinado grupo. Para avaliar a relevância é necessário identificar a relevância associada aos conteúdos através da Expressão 3.1.

$$R_u \rightarrow G = \frac{Sum(R_c) \in U_c}{Sum(R_c) \in AllU_c} \quad (3.2)$$

U_c representa um utilizador que tenha partilhado informações nos grupos, enquanto que $AllU_c$ representa todos os utilizadores que partilharam informações no grupo. A função Sum representa a soma de todos os valores de relevância de conteúdos encontrados tanto para o utilizador (U_c) como para todos os utilizadores. Esta expressão permite que seja analisado e caracterizado o comportamento dos utilizadores dentro do grupo. R_c representa a relevância dos conteúdos produzidos por um utilizador através da Expressão 3.1.

Desta forma, podem ser identificados quais os utilizadores que contribuem de uma forma mais enriquecedora na actividade do grupo, partilhando informação relevante ao grupo e os que produzem informação pouco relevante. O facto de se mostrar aos utilizadores a popularidade dos seus contributos, pode ajudar os utilizadores a melhorar a qualidade das suas partilhas e também reduzir a quantidade de informação pouco relevante partilhada. Não foram colocados limites mínimos para considerar um conteúdo relevante. Todos os conteúdos partilhados de um certo utilizador são contabilizados desde que algum indicador de interesse lhes seja atribuído.

Os utilizadores podem demonstrar interesse por um grupo, gostando ou comentando os conteúdos partilhados. O facto de os utilizadores partilharem conteúdos num determinado grupo, também demonstra um certo interesse por parte dos utilizadores por esse grupo. Após o cálculo da relevância de conteúdos e dos utilizadores para os grupos fará sentido também pensar na relevância do próprio grupo para o utilizador, ou seja identificar que grupos são importantes para o utilizador.

3.5 Relevância de um grupo para o utilizador

A avaliação da relevância de um grupo para cada um dos seus membros depende não só da avaliação da relevância da informação que é mantida no contexto do grupo, como também pode depender da avaliação das forças de ligação entre membros dos grupos. Nesta abordagem não se terão em conta as forças de ligação entre utilizadores para podermos justificar a manutenção de um grupo pois iremos focar-nos apenas na relevância dos conteúdos partilhados nos grupos e na importância que os utilizadores darão a essa informação. A caracterização da relevância de um grupo para um utilizador, irá permitir avaliar o interesse na sua permanência nesse grupo. Por outro lado, a informação partilhada e gerida por um grupo, pode ser caracterizada em termos da sua relevância relativa (popularidade), em função de critérios de utilizadores individuais dependendo do interesse individual de um dado membro. Pode também ser avaliada em função da utilidade ou interesse nessa informação, tal como seja manifestado pelo colectivo dos restantes membros do grupo. Esta avaliação vai permitir que os utilizadores analisem e seleccionem quais os grupos em que efectivamente querem participar. Para isso são avaliados não só os conteúdos partilhados pelo utilizador no grupo que têm relevância mas também outros indicadores de interesse que o utilizador demonstrou sobre conteúdos de outros utilizadores, como os gostos, comentários ou partilha de conteúdos. Também é tido em conta que tipo de actividade existe no grupo, ou seja, se existe informação a ser partilhada de uma forma frequente ou não. Este factor é importante pois normalmente os utilizadores não estão interessados em fazer parte de grupos inactivos, ainda que isto varie, consoante o propósito de criação de um grupo. Grupos criados para festivais anuais, por exemplo terão a sua actividade mais reduzida durante o ano aumentando a actividade junto ao festival. Assim, para todos os conteúdos partilhados no grupo são analisados vários componentes como podemos ver na Tabela 3.3.

Componentes da relevância de grupos
N_{cUi} : Soma do número de comentários realizados por um utilizador dentro do grupo.
N_{lUi} : Soma do número de gostos realizados por um utilizador dentro do grupo.
N_{sUi} : Soma do número de partilhas a um conteúdo realizados por um utilizador dentro do grupo.
N_c : Soma do número de comentários realizados por todos os utilizadores dentro do grupo.
N_l : Soma do número de gostos realizados por todos os utilizadores dentro do grupo.
N_s : Soma do número de partilhas a um conteúdo realizados por todos os utilizadores dentro do grupo.

Tabela 3.3: Componentes da relevância de grupos para o utilizador abrangendo todos os conteúdos partilhados

O número de conteúdos a serem analisados irá depender do *site* de rede social usado, pois cada um deles têm políticas diferentes de privacidade e organizam a informação de forma diferente.

É também contabilizada nesta nova Expressão 3.3, a relevância dos conteúdos partilhados pelo utilizador no grupo calculados através da Expressão 3.2 $R_u \rightarrow G$. A expressão seguidamente apresentada define então a relevância de um grupo para um utilizador U_i .

$$R_G \rightarrow U_i = \left(\frac{N_{lUi}}{N_l} * W_{lg}\right) + \left(\frac{N_{sUi}}{N_s} * W_{sg}\right) + \left(\frac{N_{cUi}}{N_c} * W_{cg}\right) + (R_{ui} \rightarrow G * W_{co}) \quad (3.3)$$

Ao utilizarmos a razão entre o número total e o número que o utilizador de um determinado indicador de interesse colocou, estamos a ter uma razão que dá mais valor quando são poucos utilizadores a participar no grupo. Xie [XLC12] usou uma fórmula semelhante para calcular o peso de anotações que os utilizadores davam a um determinado produto. W_{cg}, W_{sg}, W_{lg} , e W_{co} são, respectivamente, os pesos que afectam as partilhas de endereços (*share*), os comentários, os gostos e o conteúdo partilhado. Esta expressão necessita, no entanto, de ser complementada com componentes adicionais. Para isso considera-se um factor temporal para que se detecte quando foi a última vez que o utilizador esteve no grupo, ou seja, que leu ou partilhou algo no grupo (T_g) em termos do número de dias decorridos. Outro factor importante, é ter em conta se esse grupo já teve alguma relevância no passado para esse utilizador. Este factor é tido em conta para beneficiar grupos que já tiveram um historial de relevância positiva para um utilizador.

Na expressão anterior só será contabilizada a relevância que se reporte a um passado de menos de 62 dias, que representam dois meses. Também o factor temporal de quando foi a última visita de cada utilizador ao grupo só é contabilizado se for há menos de 2 meses. Segundo Shein [SPUP02] concluí-se que um dado é frequentemente considerado interessante em um pequeno espaço de tempo começando com o seu auge de popularidade quando é publicado. Neste caso com a contabilização temporal da relevância, tenta-se dar mais importância ao que é mais recente seguindo um pouco a mesma ideia.

Estes novos componentes vão ser descritos através da Tabela 3.4.

Propriedades da relevância do grupo para o utilizador
T_g : Período de tempo desde que o utilizador visitou o grupo em dias.
W_t : Peso do factor tempo.
W_o : Peso das relevâncias passadas de um grupo para o utilizador.

Tabela 3.4: Propriedades da relevância do grupo para o utilizador

$$R_G \rightarrow U_i = \left(\log\left(\frac{62}{T_g}\right) * W_t\right) + \left(\frac{N_{Wi}}{N_l} * W_{lg}\right) + \left(\frac{N_{sUi}}{N_s} * W_{sg}\right) + \left(\frac{N_{cUi}}{N_c} * W_{cg}\right) + \left(R_{Ui} \rightarrow G * W_{co}\right) + Sum(R_G \rightarrow U_i \frac{\log\left(\frac{62}{R_o}\right)}{\log(62)}) * W_o \quad (3.4)$$

Sendo que $W_{cg} + W_{lg} + W_o + W_t + W_{co} = 1$

O uso do logaritmo deve-se ao facto de se querer dar um efeito de que as relevâncias mais antigas, terão menos peso na expressão. Para isso, dependendo do número de dias da última relevância, essa relevância vai tendo mais ou menos valor.

Baseado nesta avaliação aos grupos, os utilizadores podem observar o comportamento dos seus grupos e verificar quais os grupos que realmente participam e em que têm interesse em receber informação.

3.6 Idealização de uma aplicação no *Facebook*

Depois de vários conceitos considerados, definiu-se uma aplicação, com vista à análise da relevância da informação produzida pelos utilizadores no contexto dos seus grupos e também como forma de avaliar a sua participação. Pretende-se relacionar o perfil dos grupos com a informação produzida pelos utilizadores membros desses grupos. Como já foi referido anteriormente os grupos implícitos resultam de uma agregação automática de indivíduos e conteúdos por eles partilhados que têm por base padrões de semelhança entre os seus perfis, gostos, interesses e interações. Ao assumir que a formação desses grupos é automática, ou seja, não envolve uma acção explícita por parte do utilizador, podem ser levantadas algumas questões relativamente à "utilidade" desses grupos para os seus participantes e vice-versa. Surge assim a necessidade de ter algum mecanismo de validação da relevância dos membros para o grupo e também até que ponto esse grupo e as informações nele partilhadas, são relevantes para o utilizador.

Com este objectivo em mente estruturou-se uma aplicação que permita avaliar o grau de utilidade dos utilizadores nos seus grupos, com base na informação por estes produzida e também com base numa avaliação simples dos restantes membros. Como já foi referido, o ambiente escolhido para o desenvolvimento da aplicação foi a plataforma *Facebook* não só pelo facto do seu grande número de utilizadores como também pelo facto de permitir obter-se bastante informação relativa aos grupos. Em termos de informação

do utilizador, vamos analisar, numa primeira fase, o conjunto de conteúdos (*posts*) feitos em páginas do *Facebook* pelo utilizador. Com base nessa análise serão extraídos os tópicos principais encontrados em cada *post*. De acordo com esses tópicos os *post* serão publicados ou não nos grupos cujo perfil possua essas características. No fundo pretende-se disseminar a informação do utilizador pelos seus grupos que tenham temas semelhantes ao extraídos dos *posts* do utilizador, seja para grupos implícitos ou explícitos, de acordo com o perfil do grupo. A informação original produzida é submetida à análise de uma ferramenta de análise de conteúdos *Alchemy* [Alc12b], que extrai uma lista de tópicos, os quais são depois utilizados para seleccionar as informações a disseminar face aos perfis dos grupos.

Numa segunda fase é efectuada a avaliação de relevância dessa informação no contexto de grupo com base no facto dos restantes membros terem mostrado explicitamente o seu interesse por essa informação, calculada através dos indicadores de interesse. Com base nessa avaliação e na quantidade de informação produzida pelo utilizador para o grupo, é sugerida a sua remoção ou manutenção nos respectivos grupos. Esta aplicação permite relacionar a origem da informação, ou seja, o utilizador, com os elementos que a consomem, ou seja, os membros do grupo. Esta aplicação pretende constituir uma ferramenta com um interesse especial na avaliação da utilidade/visibilidade de grupos não somente na plataforma *Facebook*. Note-se que embora esta aplicação seja desenvolvida sobre o *Facebook*, os resultados obtidos poderão vir a ser aplicados no contexto de outras plataformas de redes sociais. Esta aplicação está idealizada para ser implementada na linguagem *PHP* [ABC⁺12] pois irá facilitar a integração com as *APIs* do *Facebook* e com o *Alchemy*, como será detalhadamente explicado no Capítulo 5.

3.7 Instanciação do modelo no contexto do *Facebook*

Cada *site* de rede social tem as suas próprias especificações, formas organizacionais e distintas formas de acesso aos dados. Há certos dados que numa rede social se consegue obter e que noutra é completamente diferente, verificando-se que as *APIs* estão estruturadas de uma forma diferente. Neste caso e usando a rede social *Facebook*, foram feitas algumas reestruturações devido ao facto de existirem certas informações que não são possíveis de obter conforme se descreve nas secções seguintes.

3.7.1 Partilha de conteúdos ou endereços - *Share*

O botão de partilha de conteúdos, *share*, tem sido cada vez menos utilizado em prol do botão de *Like* no *Facebook*. O *Facebook* está a descontinuar este conceito como se pode constatar através da *API* [Fac12a]. Apenas os endereços que o *Facebook* consegue aceder e verificar, permitem que a partilha desse *URL* seja feita. Dentro dos próprios grupos não existe a informação sobre se um dado *URL* foi partilhado e quantas vezes o foi.

Dessa forma as variáveis a considerar no cálculo da relevância de um conteúdo referem-se apenas aos componentes dos comentários e dos gostos. A Expressão 3.1, apresentada no Capítulo 3, é assim reestruturada dando origem a uma expressão sem as partilhas de endereços *URL* :

$$R_c = \left(\frac{N_c}{A_c} * W_c\right) + \left(\frac{N_l}{A_l} * W_l\right) \quad (3.5)$$

Sendo que $W_c + W_l = 1$

Tanto a Expressão 3.3 tanto a Expressão 3.4 também tiveram modificações devido à remoção do indicador de interesse relacionado com as partilhas de conteúdos. Ambas as expressões retiraram assim esse componente dando origem a uma expressão actualizada:

$$R_G \rightarrow U_i = \left(\log\left(\frac{62}{T_g}\right) * W_t\right) + \left(\frac{N_{lU_i}}{N_l} * W_{lg}\right) + \left(\frac{N_{cU_i}}{N_c} * W_{cg}\right) + (R_{U_i} \rightarrow G * W_{co}) + Sum(R_G \rightarrow U_i \frac{\log\left(\frac{62}{R_o}\right)}{\log(62)}) * W_o \quad (3.6)$$

Sendo que $W_{cg} + W_{lg} + W_o + W_t + W_{co} = 1$

O peso que cada variável W_{cg} , W_c , W_l , W_t , W_{lg} , W_{co} e W_o terá na implementação da aplicação, será discutido no Capítulo 5.

3.7.2 Variável temporal

A variável temporal T_g por definição deveria significar há quanto tempo é que um utilizador não consome nada de um determinado grupo ou seja se lê ou participa na partilha de informação num grupo. Quando no *Facebook* se acede à página dos grupos tem-se uma lista de grupos e a última vez que acedemos a esse grupo. Essa informação existe para consulta através do perfil no *Facebook* do utilizador, no entanto a extracção dessa informação para aplicações, é restrita apenas aos programadores do *Facebook*, pelo que não é possível ter acesso a ela. Para se ultrapassar este problema, em vez de se utilizar esta informação, usa-se a informação de quando foi a última vez que houve actualizações no grupo, seja um novo *post* ou um comentário. Mesmo que um utilizador consulte um grupo que já não é actualizado há mais de dois meses, a informação que lá está é a mesma e não havendo actualizações demonstra que os utilizadores não têm interesse pelo grupo. O valor desta variável T_g é representado pela Expressão 3.7, em que a variável d representa o número de dias que passaram desde a última actualização ao grupo.

$$T_g(d) = \begin{cases} 0 & \text{if } d > 62 \\ \log(d/62) & \text{if } d \leq 62 \text{ e } d > 1 \\ 1 & \text{if } d \leq 1 \end{cases} \quad (3.7)$$

A execução temporal de toda esta aplicação ocorrerá de cada vez que um utilizador entra na aplicação no *Facebook*, ou seja a sua execução necessita sempre que o utilizador a provoque. Não seria útil estar sempre a recalculá-la sempre que existisse uma actualização nos grupos analisados do utilizador, pois poderia estar a processar dados que nunca chegariam a ser visualizados pelo utilizador.

3.7.3 Número de conteúdos extraídos

Existe uma limitação por parte da extracção dos *posts* do *Facebook*. No caso de um grupo a extracção está limitada aos 25 *posts* mais actualizados que foram partilhados pelo grupo. Nesse caso os dados sobre as relevâncias dos utilizadores no grupo são calculados através dos últimos 25 *posts*, no caso de o grupo ter pelo menos esses 25.

No entanto, no caso da informação sobre os *posts* do utilizador a restrição ao nível das respostas é de 5000 respostas como se encontra ao se consultar a *API* [Lim12]. Averiguou-se que existe um número limite, não documentado na *API* mas terá a ver com a antiguidade dos *posts*. Os *posts* muito antigos só são obtidos no caso da consulta se referir a esses dados de forma temporal. No caso desta proposta apenas se analisam os últimos 60 *posts* feito pelo utilizador.

3.8 Resumo

Após a proposta de uma forma de calcular a popularidade da informação encontrada nos grupos é necessário pensar numa forma para promover o seu desenho a um caso real. Existe trabalho ainda conceptual na criação de uma aplicação que use estas fórmulas de cálculo da relevância. Esta abordagem mais conceptual terá de ser readaptada a uma rede social, para que seja possível afinar e modular toda esta proposta para uma implementação real. Através das fórmulas do cálculo da relevância de utilizadores e de grupos poder-se-ão perceber os comportamentos dentro dos grupos e perceber o que é importante para os grupos e para os utilizadores neste contexto. Num contexto de um *site* de rede social isto permitiria filtrar que informação se considera importante ou não. No Capítulo 4 descreve-se a arquitectura concebida para suportar o modelo apresentado neste capítulo e as adaptações que foram necessárias para a aplicar no caso real da aplicação desenvolvida no *Facebook*.

4

Arquitectura de suporte ao modelo

Neste capítulo, vamos proceder à descrição da arquitectura da aplicação proposta que suporte através das expressões propostas no Capítulo 3, o modelo proposto. Esta aplicação irá permitir calcular e mostrar as relevâncias das informações dentro dos grupos, tornando assim possível identificar que utilizadores participam de uma forma mais "motivadora" no grupo e também perceber quais os grupos em que cada utilizador participa com um nível de interesse mais significativo.

4.1 Organização lógica da arquitectura

Nesta secção faremos uma breve descrição de todos os componentes da aplicação idealizada para o *Facebook* e explicaremos a forma como o fluxo das operações se vai realizar. A arquitectura conceptual da aplicação está representada através da Figura 4.1 onde se pode ter uma visão geral de como os processos vão acontecer. Existem dois componentes externos à arquitectura da aplicação, responsáveis por duas funcionalidades distintas:

1. *Facebook*, modulo responsável pela extracção dos dados relativos aos perfis de utilizadores e grupos;
2. Extractor de conteúdos (*Alchemy*), responsável por extrair tópicos que identifiquem os temas dos grupos e dos conteúdos partilhados pelo utilizador.

Os componentes internos da aplicação irão interagir com os componentes externos através das chamadas às respectivas *APIs*. Existem diversos componentes internos que interagem entre si e com algumas dependências importantes e que justificam serem referidas.

Para se perceber a forma como a aplicação foi idealizada e qual será o fluxo de operações da aplicação, faz-se, para cada componente, uma breve descrição das suas interações e das funções de que é responsável.

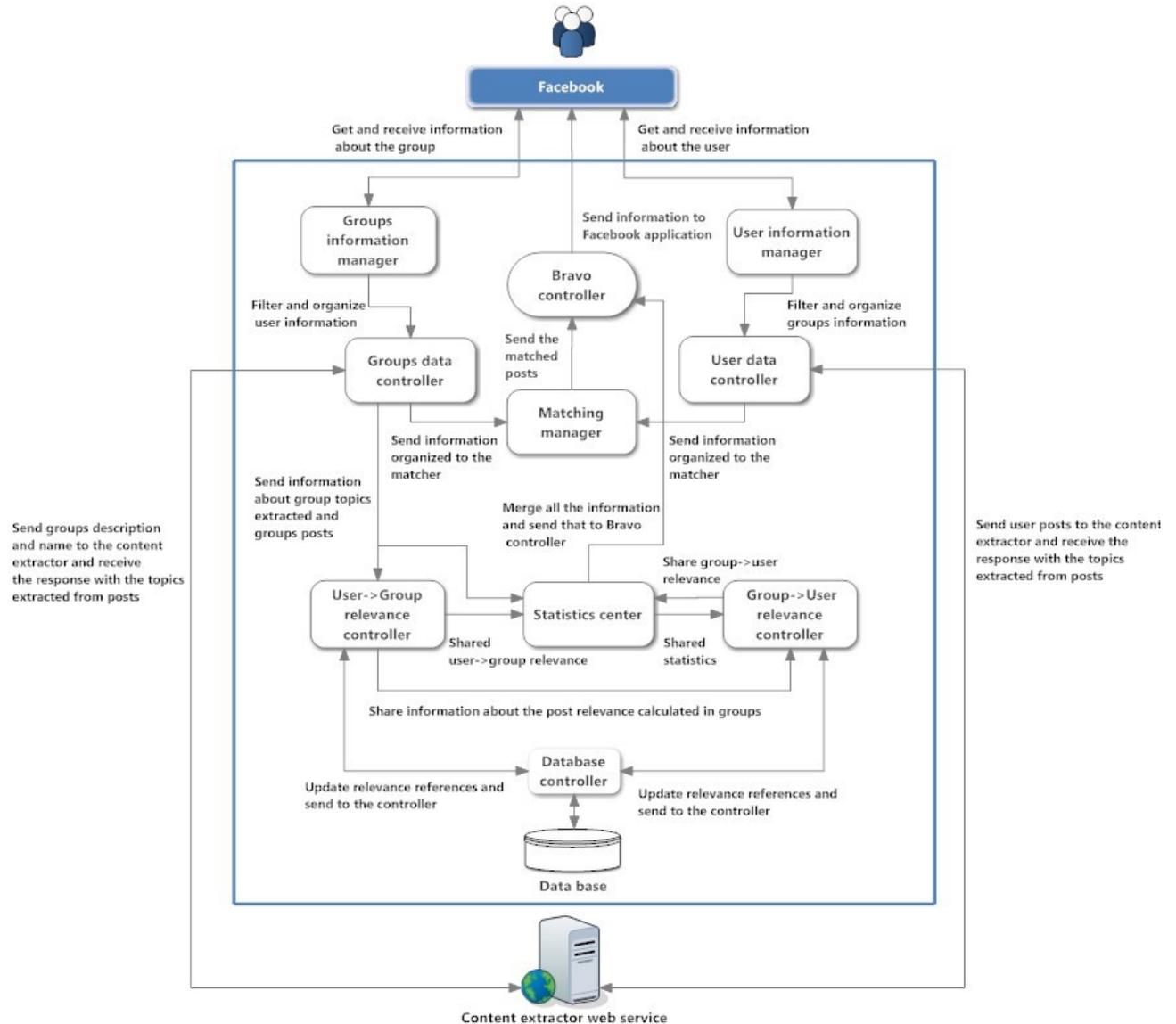


Figura 4.1: Arquitectura da aplicação no *Facebook*

4.1.1 Gestor de informação de grupos (*Groups information manager*)

Este componente é responsável por receber a informação extraída do *Facebook* sobre os grupos e organizá-la estruturadamente. Este gestor irá guardar em base de dados as informações relativas aos grupos como:

- Identificação de Grupo;
- Nome;

- Descrição;
- Número de membros.

Serão também extraídos os conteúdos (*posts*) partilhados no contexto de grupo. Estes *posts* contêm informação sobre quem partilhou uma certa informação, o número de *likes* e o número de comentários que esses *posts* obtiveram. Os conteúdos dos *posts* dos grupos não serão armazenados numa base de dados mas serão analisados e guardados em memória para serem processados por outros componentes.

4.1.2 Controlador de dados dos grupos (*Groups data controller*)

Este componente recebe dados vindos do gestor de informação dos grupos e é responsável por enviar para o extractor de conteúdos, os dados relativos ao grupo como o nome e a descrição através de uma chamada ao servidor do *AlchemyAPI*. A resposta do servidor a esta chamada será em formato *JSON [JSOto]* com as informações catalogadas identificadas pelo *Alchemy*. Este componente é também responsável por receber a resposta e verificar se existe logo à partida algum resultado que possa ser excluído, por exemplo se não se tiver detectado nenhuma categoria para o grupo, é escusado enviar a informação sobre aquele grupo para o componente que faz as correspondências que irá ser explicado adiante na Subsecção 4.1.5.

4.1.3 Gestor de informação de utilizadores (*User information manager*)

Este componente é responsável por receber a informação através do *Facebook* sobre os utilizadores e organizá-la estruturadamente. É responsável por guardar informação sobre o utilizador na base de dados com os seguintes dados :

- Identificação de utilizador;
- Nome.

Também os conteúdos dos *posts* e actividades na página do utilizador serão analisados. Os *posts* partilhados pelo utilizador (*posts*) também como os dos grupos não serão guardados em base de dados mas serão analisados e processados por outros componentes. Será processada também a data da última publicação, e este campo será utilizado para mostrar na aplicação a data da sua última publicação de forma a também incentivar a partilha de informação.

4.1.4 Controlador de dados do utilizador (*User data controller*)

Inicialmente este componente recebe dados do gestor de informação do utilizador e é responsável por enviar os dados com os *posts* para o extractor de conteúdos, através de uma chamada ao servidor do *AlchemyAPI*. Este é também responsável por receber a resposta e verificar se existe logo à partida algum resultado que possa ser excluído.

4.1.5 Controlador de correspondências (*Match controller*)

Este componente recebe a informação dos outros componentes sobre os utilizadores e sobre os grupos e faz a correspondência (*matching*). Este componente decide que *posts* é que têm as propriedades necessárias para serem partilhados para os grupos. As propriedades necessárias serão as que têm informação suficiente para existir um *matching* com alguns dos tópicos dos grupos. O resultado é guardado para depois toda a informação ser partilhada no *Facebook* através do componente que trabalha como um controlo de qualidade, o *Bravo Boss*.

4.1.6 Controlador de relevância de um utilizador nos grupos $U \rightarrow G$ (*User Group relevance controller*)

Este componente é responsável por calcular através dos dados fornecidos sobre os grupos, a relevância de um certo utilizador no grupo através das expressões mostradas no Capítulo 3. Os dados são calculados desde que haja uma actualização no grupo e guardados na base de dados. No caso de não ter existido nenhuma actualização no grupo é apenas necessário consultar a base dados e verificar a relevância que está guardada. Os dados são enviados para *Bravo Boss* que recebe os dados e os processa de forma a serem enviados para o *Facebook*.

4.1.7 Controlador de relevância de um grupo para o utilizador $G \rightarrow U$ (*Group User relevance controller*)

Este componente é responsável por calcular através dos dados fornecidos sobre os grupos, a relevância de um certo grupo para um certo utilizador seguindo as expressões do Capítulo 3. Através de uma base de dados que guarda os dados históricos sobre a relevância dos grupos, é possível guardar os dados relativos aos históricos e às estatísticas anteriores. Recebe informação sobre há quanto tempo é que os membros de um grupo já não partilham informação. Este componente também recebe informações acerca da relevância de um utilizador para um determinado grupo para poder concluir os seus dados e facilitar a implementações das expressões, já que nesta formula se considera qual a relevância dos *posts* do utilizador num determinado grupo. Os dados são enviados para *Bravo Boss* que recebe os dados e prepara tudo para serem enviados para o *Facebook*.

4.1.8 Controlador de estatísticas (*Statistics center*)

Este componente é responsável por calcular através dos dados fornecidos sobre os grupos e sobre o utilizador as suas estatísticas como por exemplo o número de membros de cada grupo, o número de *posts* extraídos, o número de *likes*, o número de utilizadores diferentes a comentar e a partilhar informação entre outros. Basicamente, este componente traduz os dados vindos do *Facebook* em dados estatísticos que serão analisados mais tarde.

4.1.9 Bravo Boss (*Bravo controller*)

Este componente é um “agilizador” dos resultados obtidos pelos outros componentes e trabalha como se fosse um controlador da qualidade dos dados. Recebe os dados dos outros componentes e interpreta-os de forma a poderem ser enviados para a interface da aplicação no *Facebook*.

4.1.10 Controlador da base de dados (*Database controller*)

Este componente é responsável pelas actualizações dos dados da relevância de grupos e utilizadores na base de dados. É também através deste controlador que os dados anteriores da relevância são consultados e enviados para os outros componentes.

4.2 Dados obtidos através do *Facebook*

Esta secção tem como objectivo mostrar que dados se podem obter através das *APIs* do *Facebook* e que modificações e melhoramentos irão ser feitos à proposta inicial. Explicaremos como obter os dados do perfil de utilizador e por fim do perfil de grupo.

4.2.1 Perfil do utilizador

Um utilizador no *Facebook* tem associado um perfil único, composto pelas suas características e interesses, onde é possível registar informação que irá servir para se poder partilhar informação nos grupos no *Facebook*, através da sua categoria.

O *Facebook* vê um utilizador como um objecto com diferentes campos. Os campos que podem ser acedidos pelo *Facebook* através das permissões básicas são:

1. Identificador;
2. Nome;
3. Género;
4. Língua;
5. Endereço da página de perfil;
6. Nome de utilizador (*Username*).

Para se obterem outras informações mais específicas, é necessário obter as permissões por parte do utilizador (denominados *access_token*). Por exemplo, se for necessário verificar qual a data de aniversário de um utilizador é necessário requerer o acesso à propriedade *data de aniversário (birthday)*, obtida através da permissão *user_birthday*. No caso de querermos obter informação sobre as mensagens que o utilizador publicou e de uma forma pública, existem várias formas distintas através das bibliotecas do *Facebook* [Fac12c] como podemos ver na Tabela 4.1. Neste caso foi escolhido o campo */posts* pois retorna o que é

Campo	Descrição
/posts	A <i>API</i> retorna os <i>posts</i> criados pelo utilizador na sua página e na de outros incluindo todo o tipo de conteúdos como endereços, fotografias ou actualizações de estado
/feed	A <i>API</i> retorna toda a informação que pode ser vista na sua página (neste caso apenas 25 conteúdos de todo o tipo)
/statuses	A <i>API</i> retorna apenas as actualizações feitas à página do utilizador que tenham uma mensagem

Tabela 4.1: Comandos da *API* para aceder à informação sobre os *posts* do utilizador

pretendido para a aplicação. Os *posts* serão então guardados e tratados na aplicação pelo gestor de informação dos utilizadores descrito na Secção 4.1. No caso do perfil de grupo haverá mais informação importante a ser processada e guardada na aplicação conforme se discute seguidamente (Subsecção 4.2.2).

4.2.2 Perfil de grupo

Existem propriedades que são específicas dos grupos no *Facebook*. Neste caso a aplicação irá analisar os grupos que são :

1. Secretos - Não é permitido obter qualquer informação do exterior sobre o grupo;
2. Públicos - Toda a informação é pública;
3. Fechados - Apenas é possível obter informação sobre os membros mas não sobre as mensagens partilhadas no grupo;

Analisando mais detalhadamente a documentação da *API* podemos observar que praticamente é possível obter toda a informação sobre os mesmos. A informação sobre os grupos que está disponível é apresentada na Tabela 4.2 .

Grupos

Esta informação do grupo é uma informação mais estática pertencente ao perfil base

Campo	Descrição
Identificador	Identificador do grupo
Versão	Versão do grupo para o <i>Facebook</i>
Ícone	<i>Url</i> para o ícone do grupo
Criador	Quem foi o utilizador que criou o grupo
Nome	Nome do grupo
Descrição	Uma breve descrição do grupo
Link	Link do grupo
Privacidade	Qual o grau de privacidade dada ao grupo, ou seja se é publico secreto, ou fechado
Data de actualização	Data da ultima actualização que foi feita ao grupo.

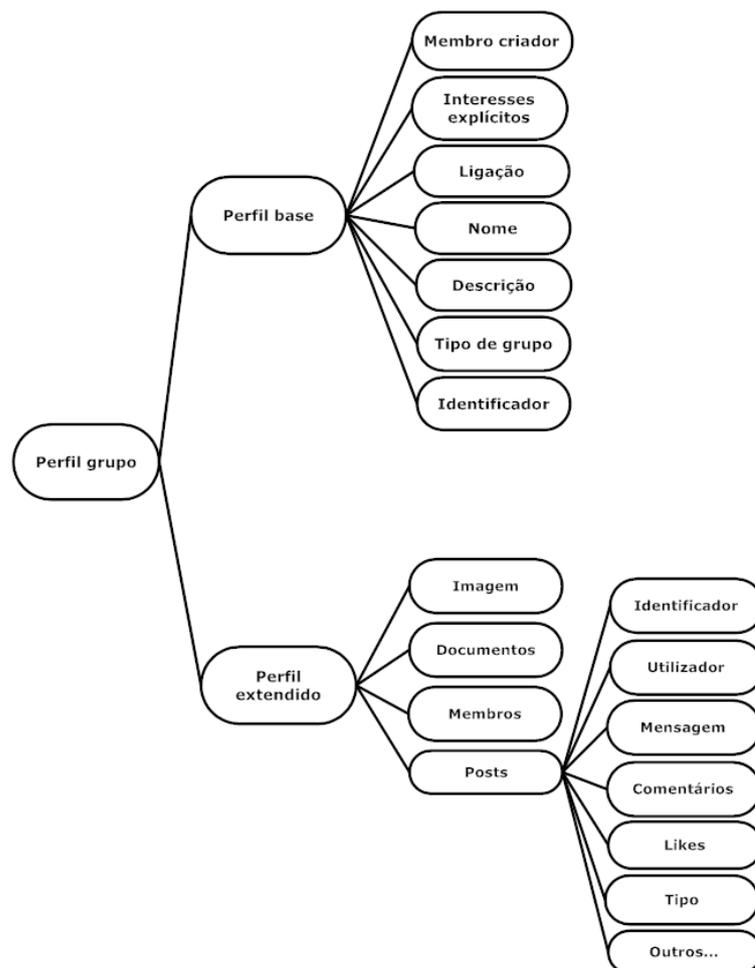
Tabela 4.2: Perfil de grupo no *Facebook*

do grupo. Em relação ao perfil estendido do grupo as informações que são possíveis de obter estão apresentadas na Tabela 4.3.

Campo
<i>Posts</i> (Array de <i>posts</i> , com tamanho máximo dos 25 últimos <i>posts</i>)
Membros (Array com os membros do grupo)
Imagem do grupo
Documentos do grupo

Tabela 4.3: Perfil de grupo estendido no *Facebook*

Ao próprio conjunto de *posts* é possível obter algumas informações como: estatísticas sobre por exemplo quem é que enviou aquele *Post*, quantos *likes* é que este tem, quem foram as pessoas que colocaram os *likes*, que comentários existem aos *posts* e respectivos utilizadores e *likes*. Com base nas informações que são possíveis obter sobre os grupos, existe uma readaptação do perfil de grupo proposto no Capítulo 3. De um perfil de grupo mais generalista, vamos propor um mais específico, com os dados do *Facebook* como podemos ver na Figura 4.2.

Figura 4.2: Perfil de grupo no *Facebook*

4.3 Dados obtidos através da API do *Alchemy*

A aplicação utilizou para analisar os conteúdos dos *posts* e das informações do grupo, o *web server* do *Alchemy*. A API do *Alchemy* [Alc12b] assenta em modelos de processamento baseados em linguagem natural e algoritmos de aprendizagem automática para analisar conteúdos e extrai meta-dados sobre informação como locais, pessoas, relações, autores, etc. A API foi concebida para análise de conteúdos na Internet através de um endereço de uma página (*URL*), através de *HTML* ou até conteúdo em texto.

Este serviço permite até 1000 chamadas ao seu servidor por dia, através de uma chave gratuita. Porém é possível, através de um pedido de uma chave académica, ter até 30000 chamadas diárias para efeitos de investigação. Devido ao número de chamadas por grupo que seria suposto fazer na aplicação, foi necessário proceder a esse pedido de uma chave académica. Através do *Alchemy* e enviando um *HTML*, um endereço ou texto é possível obter vários componentes como:

- Entidades - Identificação de pessoas, empresas, organizações cidades entre outras entidades que podem ser descobertas através da análise de conteúdos;
- Anotações - Identificação de anotações (*tags*) associadas a conteúdos, baseado nas reacções comuns humanas;
- Palavras-chave - Identificação de tópicos ou palavras chave em conteúdos;
- Relações - Identifica factos, relações e sub acções relacionas com dados entidades;
- Categorização - Identifica através dos conteúdos uma determinada categoria. Por exemplo, "crise económica na França", pertence à categoria de economia;
- Língua - Através de determinado conteúdo, consegue identificar em que língua esse está escrito.

Existem outras funcionalidades da API que não serão referidas por não serem utilizadas na implementação desta aplicação. A forma como estas funcionalidades podem ser executadas, por se tratar de algo bastante específico da implementação será explicada no Capítulo 5. É possível obter vários tipos de respostas do servidor às chamadas da API como um *JSON*, *XML* ou *RDF* [RDF12], entre outros já referidos no Capítulo 2. A arquitectura da aplicação que irá receber informação do servidor do *Alchemy*, vai ter por base as respostas em *JSON*, pois facilmente é integrado com a linguagem *PHP*, usada para a implementação da aplicação. Este *JSON* para além de ter associado valores como as categorias, relações, entidades, etc, terá também um valor representado como o grau de certeza de um determinado valor, por exemplo, grau de certeza que uma notícia sobre computadores, pertence à categoria de computadores. Esse grau de certeza poderá indicar se um determinado conteúdo será ou não avaliado, caso o grau de certeza seja demasiado baixo. No entanto se ao analisar os conteúdos produzidos neste caso pelo

utilizador no *Facebook*, se determinar que existem sempre poucos valores de correspondência, pode-se ignorar este grau de certeza.

4.4 Aplicações no *Facebook*

Após a análise de como os dados serão enviados do *Alchemy* serão passados para a aplicação, apresentaremos as aplicações no *Facebook*. O *Facebook* permite que seja possível implementar aplicações com diversos intuitos como jogos [Zin12] [Sim12], meios publicitários [Men12], entre outros que podem ser executados através da interface do *Facebook* [Fac12d]. Existem vários componentes das aplicações no *Facebook* como a *Canvas Page*¹, os canais sociais, que aumentam a população de utilizadores que usam a aplicação e as suas estatísticas obtidas através do *Analytics*[Fac12e]. Estas aplicações usam várias linguagens como *PHP* ou *Javascript* para poderem aceder aos dados que estão no *Facebook*. Estas aplicações funcionam em servidores independentes que utilizam o *Facebook* como uma interface. Quando se cria uma aplicação no *Facebook*, para além de informações como o nome da aplicação, temos de colocar os endereços de redireccionamento. Para proceder à criação de uma aplicação é necessário respeitar alguns passos específicos. É necessário inscrever a aplicação no centro de aplicações do *Facebook* com alguns dados da mesma e esperar pela sua confirmação que é enviada através de uma mensagem escrita para o telefone do programador confirmando a autenticidade do mesmo. É necessário colocar o nome da aplicação e o *link* da página onde a aplicação executa. É ainda necessário para além de um link comum *HTTP*, colocar um link para aplicação com ligação segura ou seja *HTTPS* [Htt12]. É necessário colocar este campo, com o endereço de ligação segura, para os utilizadores que usam o *Facebook* com o modo ligação segura activa [Fac12b] que é bastante frequente. Quando um utilizador instala uma aplicação existe sempre a necessidade de requerer autorização para obter informação do utilizador, quer seja informação base quer seja informação mais específica como os seus grupos, a sua localização ou o seu e-mail. Por omissão só é obtida informação base, sendo no entanto possível obter mais informação desde que tal seja especificado no pedido. Esta informação só é acedida no caso de o utilizador aceitar autorizar que a sua informação seja enviada à aplicação através de janela como podemos ver na Figura 4.4. Esta janela, conhecida como *AuthDialog*, mostra a informação das permissões de que a aplicação a instalar vai necessitar. É importante referir que para a autenticação dos utilizadores ser feita através das credenciais do *Facebook*, é necessário escolher correctamente quais as permissões que vão ser necessárias para a implementação. Essas permissões podem ser definidas na própria aplicação que cria um *token* que serve como uma chave que contém as permissões codificadas necessárias para usar a *API* do *Facebook*.

É através das aplicações do *Facebook* que foram descritas que a aplicação proposta irá ser executada e mostrada aos utilizadores. Uma descrição mais detalhada da implementação da aplicação é apresentada no Capítulo 5.

¹Canvas Page - Página onde a aplicação vai executar.

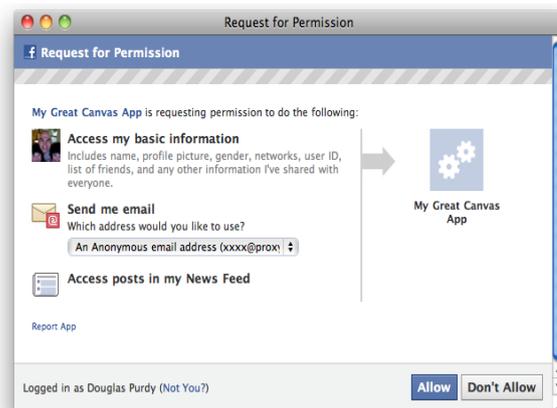


Figura 4.3: Janela do *Facebook* a pedir autorização para a instalação da aplicação.

4.5 Resumo

Algumas modificações ao modelo proposto foram feitas devido à necessidade da sua adaptação ao caso real do *site* de rede social *Facebook*. As aplicações no *Facebook* têm algumas propriedades próprias e específicas assim como a própria *API* do *Facebook*. Poderão existir mais modificações que terão em conta já a parte da implementação numa linguagem e também as modificações necessárias devido ao facto de existirem componentes externos à aplicação. No próximo capítulo descrevem-se as várias camadas da aplicação e os componentes da sua implementação.

5

Implementação

Neste capítulo iremos abordar a implementação do modelo proposto e descrever as principais decisões de concepção e de implementação. O nome escolhido para a aplicação foi *GIA*, *Group Information Agregator*, pois como esta aplicação, na prática irá agregar e analisar informação para os grupos ou nos grupos, este nome surgiu como adequado. Esta aplicação irá agregar dados com informação estatística sobre a relevância/popularidade tanto do utilizador como dos seus grupos, como irá disponibilizar uma forma de partilhar informação do utilizador para os grupos consoante o tema do grupo. Na Figura 5.1 apresentamos a arquitectura geral da aplicação *GIA*, onde se encontram os componentes que compõem a sua arquitectura interna.

O desenvolvimento desta aplicação baseia-se num modelo arquitectural de software chamado *MVC*. O desenvolvimento desta aplicação baseou-se no padrão de desenvolvimento *MVC* (*Model-View-Controller*), o que vai permitir que exista uma divisão entre a parte dos modelos de dados (*Model*), a parte de controlo do processamento e execução (*Controller*) e por fim a visualização dos dados (*View*) [KP88]. O modelo *MVC* é um padrão de arquitectura que usa 3 diferentes componentes que têm funções completamente distintas. Os modelos estruturam os dados e respondem a pedidos das vistas da aplicação e a instruções pedidas pelos controladores. Os controladores recebem os pedidos que são efectuados a partir das vistas e executam de forma a processarem a informação a ser enviada para os modelos. Basicamente os controladores fazem a parte do processamento da informação. As vistas são os componentes que compõe a interface e é a partir destas que se pode fazer pedidos de processamento aos controladores ou de dados aos modelos da aplicação. Na Figura 5.2 podemos ver a estrutura do modelo *MVC*, onde percebemos que todos os componentes podem comunicar entre si, o que difere da arquitectura das 3 camadas distintas. O facto de utilizarmos este modelo de arquitectura permite não só

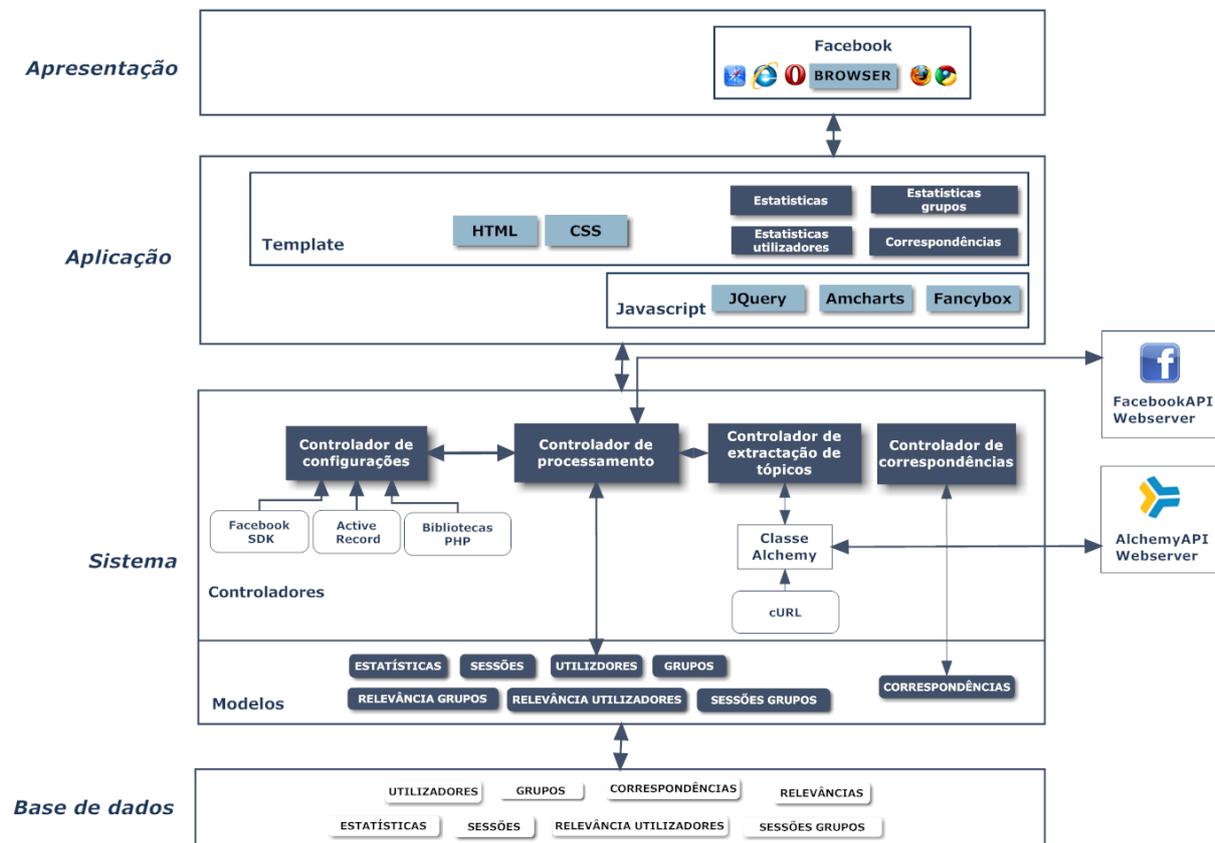


Figura 5.1: Arquitectura interna da aplicação GIA

que haja uma reutilização de código, mas também extensibilidade e organização das diversas funcionalidades [CG08]. A arquitectura foi organizada em 4 camadas distintas, cada uma com o seu objectivo distinto, que irá ser explicitado mais especificamente nas secções seguintes.

5.1 Camada de base de dados

A camada de base de dados é responsável pela organização dos dados, que serão guardados em sistema sob controlo da aplicação. Nesta secção descreve-se a estrutura da base de dados escolhida de forma a melhor representar e guardar informação da aplicação. A base de dados relacional baseia-se em *SQL*, especificamente em *MySQL* [MyS12b]. O sistema de armazenamento das tabelas escolhido, foi *InnoDB* [Inn12], pois quando necessário fazer uma inserção à base de dados este permite que apenas uma linha da tabela seja bloqueada (permitindo uma maior concorrência de transacções) e não a tabela toda como acontece com *MyISAM* [MyI12]. No caso do sistema de armazenamento desta aplicação a escolha de *InnoDB* mostrou-se adequada, pois permite-se que existam casos de vários utilizadores a aceder ao mesmo tempo e dessa forma tem de haver um controlo aos bloqueios à base de dados. Dessa forma evita-se que se perca alguma informação ou se observe informação inconsistente.

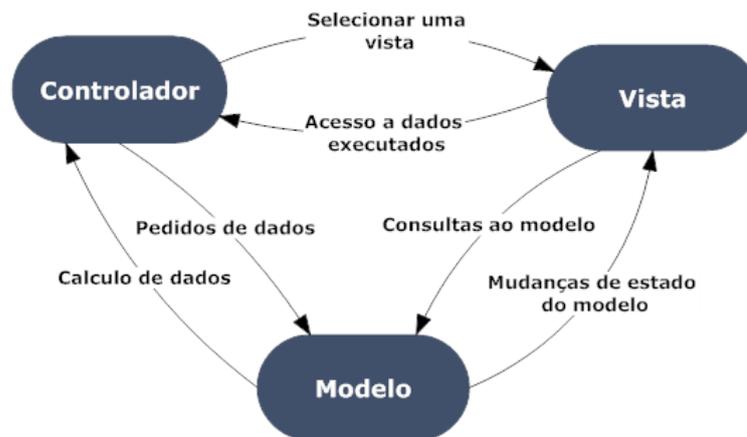


Figura 5.2: Arquitectura do Model-View-Control

5.1.1 Estrutura da base de dados

A base de dados é composta por 8 diferentes tabelas para se poder armazenar informação relativamente às estatísticas da relevância e sobre a partilha de informação nos grupos do *Facebook*. No entanto a maior parte da informação guardada encontra-se no âmbito das estatísticas do conteúdo que é partilhado dentro do grupo por utilizadores. Não foi necessário proceder ao arquivamento da própria informação partilhada (*posts*) pois iríamos estar a guardar informação que era uma réplica da que já existe no *Facebook* e que pode ser facilmente acedido através das *APIs* do *Facebook*. As tabelas criadas têm associados modelos de dados que serão descritos posteriormente na camada de sistema na Secção 5.2.

5.1.2 Desenho da base de dados

O desenho da base de dados foi construído através das necessidades da aplicação e da informação que era necessário guardar. Não é guardada informação que torne a base de dados um espelho do *Facebook* mas sim apenas a informação mais dedicada às estatísticas. A estrutura da base de dados pode ser vista na Figura 5.3. Das diferentes tabelas é possível referir algumas observações da necessidade de as mesmas serem criadas. É através da sessão de um determinado utilizador que se obtêm os dados relativos à data em que as estatísticas foram calculadas e quem foi o utilizador. Se esta variável que identifica a sessão não existisse não seria possível consultar a informação sobre as estatísticas mais recentes pois não havia uma relação entre as 3 estatísticas diferentes calculadas. Por este motivo procedeu-se à ideia das sessões. Na tabela das relevâncias dos utilizadores (*Statusergroups*) existe um campo utilizador (*user*) que não referente a um utilizador do sistema mas sim presente num grupo. Optou-se por apenas guardar em tabela os utilizadores que acederam a esta aplicação. Esse campo é o que identifica que um determinado utilizador teve relevância dentro desse grupo, sendo essa informação necessária para poderem ser mostradas as importâncias dos utilizadores nos grupos. Mais informações sobre as

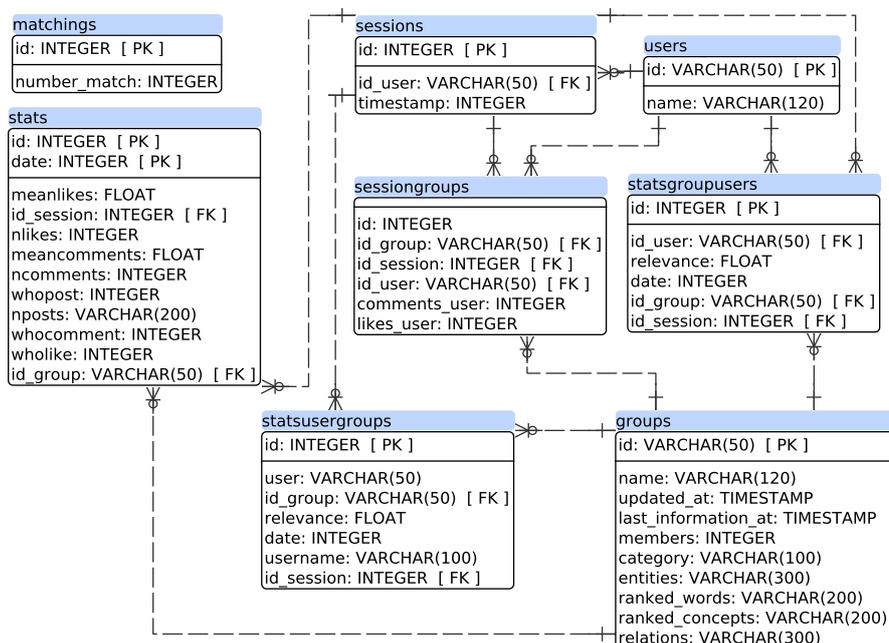


Figura 5.3: Estrutura da base de dados

características destas tabelas são descritas conjuntamente com os modelos na Subsecção 5.2.1.

5.1.3 Informação dos grupos na base de dados

Quando um novo grupo é acessado pela execução da aplicação, as informações relativas ao grupo são actualizadas, na base de dados. Na base de dados os campos seguintes são preenchidos a partir não só da informação vinda do *Facebook* mas também a partir da informação recebida através *Alchemy*.

- Nome;
- Descrição;
- Relações;
- Categoria do Grupo;
- Palavras chave;
- Conceitos;
- Entidades.

Para que estes campos sejam preenchidos é necessário enviar para o *Alchemy* o nome do grupo e a descrição do mesmo. Estes campos podem ser ou não preenchidos dependendo da informação extraída dos componentes externos. Por exemplo, é comum um

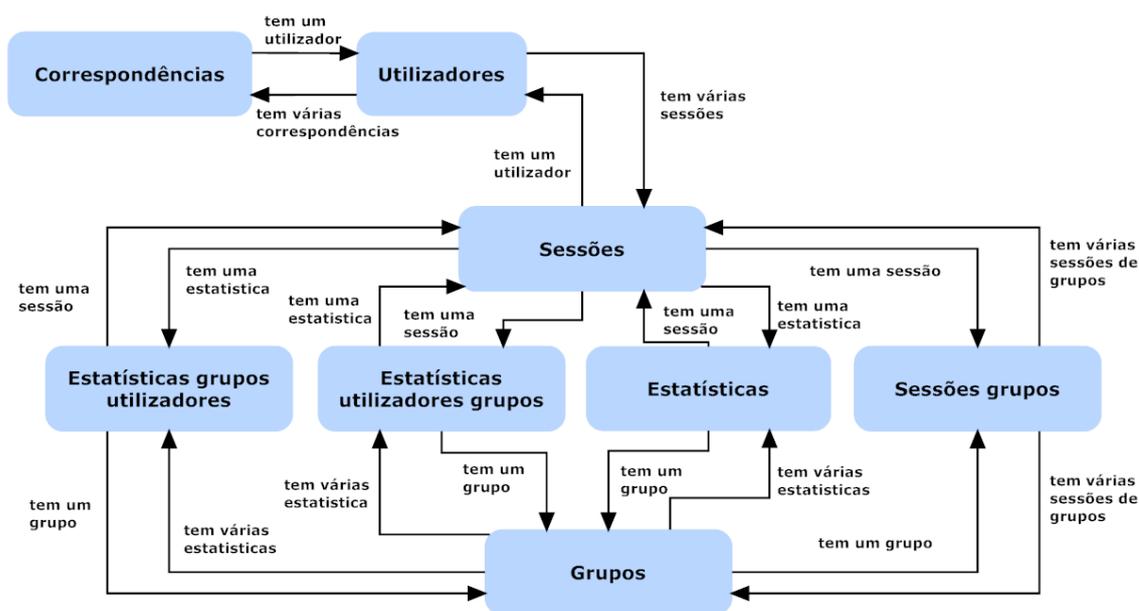


Figura 5.4: Relações entre as tabelas

grupo no *Facebook* não ter uma descrição ou a descrição ser igual ao nome. O facto de isto acontecer não permite muitas vezes que o *Alchemy* consiga obter toda a informação necessária para conseguir extrair tópicos como as categorias, entidades, relações, etc. Optou-se por guardar na base de dados esta informação para evitar realizar chamadas a *API* do *Alchemy*, sem que tenha havido qualquer actualização aos dados dos grupos. Através desta opção evitou-se desperdiçar chamadas diárias sem que fosse necessário e aumentar a velocidade de processamento da aplicação.

Quando é possível detectar na informação sobre os grupos vinda do *Facebook*, as propriedades do *Alchemy* ou seja, categorias, entidades, relações, anotações e conceitos, essas informações são preenchidas com a informação vinda do *Alchemy*. Quando existe informação suficiente na descrição, mas não é detectado qual o âmbito do grupo, o *Alchemy* classifica esse grupo com a categoria desconhecido. Quando não há informação significativa o controlador não executa as funções do *Alchemy* e na base de dados apenas fica guardada a informação relativa ao nome, membros e data da última actualização. Para controlar estas operações de extracção e correspondências entre informação de utilizadores e de grupos, existe o controlador de extracção de tópicos, apresentado na Subsecção 5.2.2.7.

5.2 Camada de sistema

Esta camada de sistema tem como principal objectivo não só coordenar as informações do contexto de utilização da aplicação, camada de apresentação e camada da base de dados, como também a coordenação da informação que vem dos componentes externos (*Facebook* e o extractor de conteúdos *Alchemy*). Para a camada que lida com a parte da

implementação do lado do servidor, decidiu-se utilizar a linguagem *PHP* como base da aplicação. Alguns componentes externos e algumas bibliotecas serão descritas na secção seguinte, na qual se apresenta uma descrição mais pormenorizada da camada de sistema. Esta camada está dividida em duas subcamadas: camada de modelos onde se encontram os modelos de dados; camada de controladores onde se encontram os controladores da aplicação.

5.2.1 Camada de modelos

A camada que define as estruturas dos dados usa modelos na sua implementação através do padrão *Active Record* [Fow03]. Este padrão foi escolhido para esta aplicação de forma a reduzir a complexidade da própria implementação e a interface entre a base de dados e a aplicação pois usa o conceito de mapeamento relacional da base de dados. O *Active Record* é um padrão arquitectural utilizado em software de armazenamento de informação em base de dados, em que um tuplo representa uma instanciação de um objecto de uma tabela relacional. As tabelas na base de dados são representadas através de classes que estendem um modelo *Active Record* e os *tuplos* representados como objectos da própria aplicação. A inserção, actualização e remoção de *tuplos* é feita de uma forma fácil e evita-se desta forma que os programadores tenham preocupações com execução de primitivas, por exemplo do SQL, pois existirá uma interface de programação que irá responsabilizar-se por essa tarefa.

Os modelos neste caso são suportados por classes responsáveis por controlar os registos da base dados e por terem alguns outros métodos auxiliares que permitem ajudar no cálculo dos dados para serem guardados na base de dados.

Os modelos criados vão ser descritos nas subsecções seguintes.

5.2.1.1 Modelo Correspondências

O modelo Correspondências ou *Matchings* é responsável por guardar os dados estatísticos sobre as correspondências realizadas entre os *posts* públicos da página do utilizador e os seus grupos pessoais. Os métodos implementados neste modelo estão representados na Tabela 5.1.

5.2.1.2 Modelo Utilizadores

O modelo Utilizadores ou *Users* é responsável por guardar a informação sobre os utilizadores que entraram na aplicação e guarda informações como o seu identificador do *Facebook*. Os métodos implementados neste modelo estão representados na Tabela 5.2.

Método	Descrição
<i>addMatch(Array information)</i>	Método que cria uma nova sessão associando uma sessão às estatísticas de um grupo, através de um vector com informação
<i>updateMatch(Array information)</i>	Método que actualiza uma sessão através de um vector com informação
<i>delMatch(int id_session)</i>	Método que apaga uma sessão através de um número de identificação
<i>matchGroup(Array valuesposts, Match group)</i>	Método responsável por fazer a correspondência entre os <i>posts</i> vindos do utilizador e os dados sobre o grupo que estão na base de dados.
<i>match(Array valuesposts, Array valuesgroups)</i>	Método responsável por fazer as correspondências no caso de ser a primeira vez que existe informações sobre o grupo na base de dados.

Tabela 5.1: Métodos do Modelo Correspondências

Método	Descrição
<i>addUser(Array information)</i>	Método que cria um novo utilizador através de um vector com informação
<i>updateUser(Array information)</i>	Método que actualiza um utilizador através de um vector com informação
<i>delUser(int id_user)</i>	Método que apaga um utilizador através de um número de identificação
<i>getUser(int id_user)</i>	Método que através do número de identificação do utilizador, devolve a informação sobre o utilizador
<i>getPostUsers(Facebook facebook, String token, int id_user)</i>	Método auxiliar que recebendo uma referencia a um objecto <i>Facebook</i> , um <i>token</i> e o número de identificação do utilizador devolve os últimos 30 <i>posts</i> da página do utilizador

Tabela 5.2: Métodos do Modelo Utilizador

5.2.1.3 Modelo Estatísticas

O modelo Estatísticas ou *Stats* é responsável por guardar as estatísticas encontradas na análise das informações dos grupos. Calcula a quantidade e as médias dos indicadores de relevância que um determinado grupo tem, como os *likes* e os comentários. Dessa forma se consegue adaptar as fórmulas do cálculo da popularidade de uma informação ao tipo de grupo que se analisa, quer tenha muitos utilizadores, sejam activos ou passivos. Os métodos implementados neste modelo estão representados na Tabela 5.3.

5.2.1.4 Modelo Relevância utilizadores para grupos

O modelo Relevância Utilizadores ou *Statsusergroups* é responsável por guardar a relevância dos utilizadores para um determinado grupo. É importante que se guardem estes dados pois a fórmula do cálculo da popularidade de um utilizador tem uma componente que contém as relevâncias anteriores do utilizador. Também no caso do grupo não ter sido actualizado desde a última vez que um utilizador entrou na aplicação, o facto de se guardar esta informação permite que não seja necessário novamente fazer o cálculo das

Método	Descrição
<i>addStat(Array information)</i>	Método que cria uma nova estatística através de um vector com informação
<i>updateStat(Array information)</i>	Método que actualiza uma estatística através de um vector com informação
<i>delStat(int id_user)</i>	Método que apaga uma estatística através de um número de identificação
<i>getStat(int id_stat,int id_session)</i>	Método que devolve uma estatística através do identificador da sessão e do identificador da estatística.
<i>listStats(Array options)</i>	Método que lista às estatísticas guardadas na base de dados.

Tabela 5.3: Métodos do Modelo Estatísticas (*Stats*)

relevâncias para esse grupo. Este modelo tem um campo de utilizador (*user*), que corresponde ao número do utilizador no grupo e não o identificador do utilizador do sistema. Os métodos implementados neste modelo estão representados na Tabela 5.4.

Método	Descrição
<i>addStatUserGroup(Array information)</i>	Método que cria uma nova estatística entre utilizador e grupo através de um vector com informação
<i>updateStatUserGroup(Array information)</i>	Método que actualiza uma estatística entre utilizadores e grupo através de um vector com informação
<i>delStatUserGroup(int id_user)</i>	Método que apaga uma estatística através de um número de identificação
<i>getStat(int id_stat,int id_session)</i>	Método que devolve uma estatística através do identificador da sessão e do identificador da estatística.
<i>listStats(Array options)</i>	Método que lista as estatísticas guardadas na base de dados.
<i>aasort (Array array, String key)</i>	Método auxiliar que através de uma chave, ordena o array onde essa chave se encontra e retorna um novo array. Este método era usado para ordenar a relevância de utilizadores para ser visualizada na interface de uma forma decrescente.
<i>delallStatUserGroup()</i>	Método auxiliar que apaga todas as entradas de estatísticas da tabela <i>statusergroups</i> base de dados .

Tabela 5.4: Métodos do Modelo Estatísticas Utilizador Grupo(*StatsUserGroup*)

5.2.1.5 Modelo Grupos

O modelo Grupos ou *Groups* é responsável por identificar um grupo no contexto da aplicação *GIA* de forma unívoca, não só para tentar evitar as chamadas ao *Facebook*, mas também ao *webservice AlchemyAPI*. Como informações como o nome e como a descrição normalmente não mudam no *Facebook*, pois este só permite mudar uma vez, o facto de guardarmos essa informação em base de dados ajuda a evitar chamadas redundantes

Método	Descrição
<i>addGroup(Array information)</i>	Método que cria um novo grupo através de um vector com informação.
<i>updateGroup(Array information)</i>	Método que actualiza um grupo através de um vector com informação.
<i>delGroup(int id_session)</i>	Método que apaga um grupo através de um número de identificação
<i>getMembers(Facebook facebook,int id_group, String token)</i>	Método que devolve os membros de um determinado grupo com o identificador id.
<i>updateGroupInfoAlchemy(Array valuesgroups,int group)</i>	Método auxiliar que trata dos dados sobre o tema do grupo antes de serem efectivamente gravados na base de dados.
<i>getGroupInfo(String token,Facebook facebook,int id_user,int id_session)</i>	Método que trata toda a informação vinda do Facebook e calcula as estatísticas do grupo e dos utilizadores através de vários métodos auxiliar deste modelo e de outros.
<i>getPosts(String token,Facebook facebook,int id_group)</i>	Método que devolve os <i>posts</i> de um grupo através do seu identificador.
<i>getLikes(Array information)</i>	Método auxiliar que dado um determinado <i>post</i> , conta o número de <i>likes</i> e identifica os utilizadores do mesmo.
<i>getComments(Array information)</i>	Método auxiliar que dado um determinado <i>post</i> , conta os comentários e identifica os utilizadores do mesmo.
<i>listGroups(Array options)</i>	Método que lista os grupos.
<i>getDiferenceBetweenDates(Date date1, Date date2)</i>	Método auxiliar que dando duas datas devolve se existe diferença entre as datas.
<i>calculateStatsByPost(Array posts,Float meanlikes,Float meancomments,Float W_c,Float W_l)</i>	Método que calcula a relevância de um <i>post</i> . Este método tem em conta se existem valores nulos para não causar erros por existirem divisões por 0.
<i>statsByPost(int likes,float meanlikes, int comments, Float meancomments,Float W_c, Float W_l)</i>	Método auxiliar que calcula a relevância de um <i>post</i> através do método <i>calculateStatsByPost</i> .

Tabela 5.5: Métodos do Modelo Grupo

aos componentes externos. Esta tabela ajuda também a ter informações como o número de membros dos grupos assim facilitando encontrar alguns padrões de comportamento consoante o número de membros dos grupos. Os métodos implementados neste modelo estão representados na Tabela 5.5.

5.2.1.6 Modelo Relevância de grupos para utilizadores

O modelo relevância grupos para os utilizadores ou *Statsgroupusers* é responsável por representar a informação sobre a relevância de um grupo para os utilizadores. Através desta tabela será possível obter estatísticas sobre o número de grupos que os utilizadores não acham relevantes para estes e até sugerir a sua saída dos grupos no caso de ser nula. Os métodos implementados neste modelo estão representados na Tabela 5.6.

Método	Descrição
<i>addStatGroupUser(Array information)</i>	Método que cria uma nova estatística entre grupo e utilizador, através de um vector com informação
<i>updateStatGroupUser(Array information)</i>	Método que actualiza uma estatística entre grupo e utilizador, através de um vector com informação
<i>delStatGroupUser(int id_user)</i>	Método que apaga uma estatística através de um número de identificação
<i>getStatGroupUser(int id_stat,int id_session)</i>	Método que devolve uma estatística através do identificador da sessão e do identificador da estatística.
<i>listStats(Array options)</i>	Método que lista as estatísticas guardadas na base de dados.
<i>getDifferenceBetweenDates(Date date1,Date date2)</i>	Método auxiliar que calcula o valor da variável temporal. Consoante o número de dias desde a ultima actualização maior ou menor será esse valor. Se a ultima actualização for há mais de 2 meses o resultado é 0. No caso da ultima actualização for a menos de 24 horas o resultado é 1.
<i>findLastRelevance(Int user,Int group)</i>	Método que soma as relevâncias passadas dos utilizadores encontradas na base de dados .
<i>listStatGroupUser(Array options)</i>	Método que lista as estatísticas sobre a relevância dos grupos para os utilizadores guardados na base de dados.

Tabela 5.6: Métodos do Modelo Estatísticas Grupo Utilizador(*StatGroupUser*)

5.2.1.7 Modelo Sessões

O modelo Sessões ou *Sessions* é responsável por guardar as sessões dos utilizadores na aplicação. Irá permitir identificar qual foi a última informação extraída de um determinado grupo e qual foi o utilizador responsável por essa sessão. É importante esta informação ser guardada para que não seja necessário recalcular todos os valores, de cada vez que é executada a aplicação. Os métodos implementados neste modelo estão representados na Tabela 5.7.

Método	Descrição
<i>addSession(Array information)</i>	Método que cria uma nova sessão através de um vector com informação
<i>updateSession(Array information)</i>	Método que actualiza uma sessão através de um vector com informação
<i>delSession(int id_session)</i>	Método que apaga uma sessão através de um número de identificação

Tabela 5.7: Métodos do Modelo Sessão

5.2.1.8 Modelo Sessões de grupo

O modelo Sessões grupo ou *Sessions group* é responsável por guardar as sessões, o grupo associado aquela sessão, o utilizador e o número de comentários e de *likes* que um utilizador colocou nos conteúdos analisados daquele grupo. Esta informação é guardada para que, se a informação das relevâncias já estiver guardada em memória, não seja necessário voltar a fazer o cálculo e apenas consultar a base de dados. Os métodos implementados neste modelo estão representados na Tabela 5.8.

Método	Descrição
<i>addSessionGroup(Array information)</i>	Método que cria uma nova sessão associando uma sessão às estatísticas de um grupo, através de um vector com informação
<i>updateSessionGroup(Array information)</i>	Método que actualiza uma sessão através de um vector com informação
<i>delSessionGroup(int id_session)</i>	Método que apaga uma sessão através de um número de identificação

Tabela 5.8: Métodos do Modelo Sessão Grupo

A cada modelo está associado um conjunto de primitivas que permitem diversas operações sobre objectos do modelo como a criação, modificação ou eliminação dos vários objectos. Estes métodos podem ser invocados pelos controladores que irão executar o fluxo dos dados.

5.2.2 Camada de controladores

Esta camada é responsável por representar os controladores do sistema que vão permitir que a informação seja modificada e tratada na aplicação. Os controladores são utilizados para a gestão lógica da ferramenta. Não foi necessário implementar um controlador para cada modelo pois existem modelos que são apenas utilizados na arquitectura interna e que apenas ajudam no processamento dos dados. Os controladores considerados aqui, trataram das configurações, do processamento dos dados e da extracção de tópicos através do *Alchemy*. Nas subsecções seguintes apresentaremos os controladores de uma forma mais detalhada.

5.2.2.1 Controlador de configurações

Este componente é responsável pela inicialização das configurações do *PHP* para serem executadas na aplicação. Este controlador baseia-se num ficheiro de configurações que guarda informações como:

- Definir os dados da base de dados como informação sobre senhas e utilizador tal como informação com o endereço onde a base de dados está alojada;

- Informações como a codificação usada, neste caso é o *UTF8*¹;
- Endereço (*URL*) onde a aplicação está a ser alojada;
- Constantes utilizadas nas programação, como o número da aplicação secreto e o identificador da aplicação público.

Este componente também tem um ficheiro responsável por inicializar as variáveis do sistema e proceder à ligação da aplicação com a base de dados do sistema. Para isso este controlador precisa de bibliotecas como o *Facebook SDK* e das bibliotecas do *Active Record* para *PHP* [JFv12] para processar a ligação à base de dados e inicializar o objecto da API do Facebook como se pode ver no código presente na Listagem 5.1.

Listing 5.1: Inicialização das configurações

```
1
2
3 ActiveRecord\Config::initialize(function($cfg) {
4     global $confBD;
5     $cfg->set_model_directory(BASE_PATH.'classes');
6     $cfg->set_connections(array(
7         'development' => 'mysql://'. $confBD['user'] . ':' . $confBD['password'] . '@
8         ' . $confBD['host'] . '/' . $confBD['basedados']));
9
10    });
11
12 $app_id = APP_ID;
13 $app_secret = APP_ID_SECRET;
14 $my_url = MY_URL;
15
16 session_start();
17
18 $facebook = new Facebook( array('appId' => $app_id, 'secret' => $app_secret));
```

A biblioteca de desenvolvimento de software do *Facebook* para o *PHP* tem a base para as aplicações desenvolvidas em *PHP* poderem obter informações que estão publicadas no *Facebook*. Esta ferramenta ajuda a fazer a autenticação dos utilizadores, usa um leque de funcionalidades do lado do servidor que irão permitir o acesso ao servidor do *Facebook* que tem como suporte a *Graph API* responsável pela extracção da informação. A biblioteca permite a leitura e escrita de informações no *Facebook*, normalmente denominado por *Graph API*. Esta permite ainda uma vista consistente de todos os objectos do *Facebook* como os seus utilizadores, comentários, *likes* ou *posts*.

O uso desta biblioteca deve-se ao facto de ser a mais actual que o *Facebook* disponibiliza. Anteriormente as aplicações usavam uma *API REST* que neste momento já foi descontinuada. Para fazer o requerimento da permissão de acesso a determinados dados é preciso obter permissões (*tokens*). Estes *tokens* podem permitir obter informação

¹UTF-8 (8-bit Unicode Transformation Format) é um tipo de codificação *Unicode* de comprimento variável criado por Ken Thompson e Rob Pike que pode representar qualquer carácter universal padrão do *Unicode*, sendo também compatível com o ASCII [Pik12].

sobre o próprio utilizador como *user_about_me*, *user_activities*, *user_birthday*, *user_checkins*. Para além da informação pessoal é possível obter informação sobre os amigos através de outros *tokens* como *friends_groups*, *friends_likes*, *friends_hometown*. Existem também *tokens* especiais que permitem o acesso a informação que é disponibilizada nas *walls* dos utilizadores onde se encontram *post* de vários utilizadores e na sua própria página. A aquisição destes *tokens* é bastante importante pois, sem eles, o *Facebook* não permite a maioria da requisição de dados.

5.2.2.2 Controlador do processamento de dados

Embora o objecto *Facebook* seja inicializado no controlador de configurações é utilizado no controlador de processamento de dados para fazer a autenticação do utilizador como podemos ver no código presente na Listagem 5.2

Listing 5.2: Autenticação do utilizador

```
1
2
3 $user = $facebook -> getUser();
4
5 $params = array('scope' => 'read_stream, _friends_likes, user_groups, user_status,
6 _read_stream, user_about_me, user_location',
7 'redirect_uri' => 'https://tangis.portalvivo.net/', 'display' => 'popup');
8
9 $loginUrl = $facebook -> getLoginUrl($params);
10
11     if($user) {
12
13
14         try {
15
16 $token=$facebook->getAccessToken();
17 ...
```

Neste código é feita a autenticação do utilizador através do objecto *Facebook* e depois com os parâmetros com as permissões (*tokens*) é requerido um conjunto de caracteres que o *Facebook* entenderá como a autorização para que esse utilizador possa obter informação. Estes *tokens* deverão ser iguais aos declarados nas configurações da aplicação para não existirem complicações de haverem permissões não autorizadas.

O processamento e cálculo dos dados é realizado através dos métodos das classes que representam os modelos. No entanto este controlador é responsável pela execução dos métodos dessas classes como os das Subsecções 5.2.2.3, 5.2.2.4 e 5.2.2.5.

5.2.2.3 Cálculo da relevância de *posts*

Para calcular a relevância de um determinado *posts* é necessário chamar o método *calculateStatsByPost* da classe grupo. Como não existe nenhum modelo associado a *posts*

colocou-se este método neste modelo. Este método recebe vários parâmetros para conseguir calcular qual a importância que os utilizadores deram a um *post*. Os parâmetros são:

- *Posts* - Variável que contém a informação sobre os *posts* extraídos através do *Facebook*;
- *meanlikes* - Média de *likes* no grupo;
- *meancomments* - Média de comentários;
- W_l - Peso dos *likes*;
- W_c - Peso dos comentários.

No pseudo código presente na Listagem 5.3 iremos demonstrar como é processado o cálculo da relevância de *posts* partilhados.

Listing 5.3: Cálculo da popularidade de um *post*

```

1
2
3 for ($i=0; $i<sizeof($posts); $i++)
4 {
5     if(!isset($posts['userslikes'])&&!isset($posts['userscomments']))
6     {
7         $r=0;
8     }
9     else if(!isset($posts['userslikes'])&&isset($posts['userscomments']))
10    {
11        $r=Group::statsByPost(0,$meanlikes,count($posts['userscomments']),
12        $meancomments,$WC,$WL);
13
14    }
15    else if(!isset($posts['userscomments'])&&isset($posts['userslikes']))
16    {
17        $r=Group::statsByPost(count($posts['userslikes'],$meanlikes,0,
18        $meancomments,$WC,$WL);
19    }else
20    {
21        $r=Group::statsByPost(count($posts['userslikes'],$meanlikes,
22        count($posts['userscomments'],$meancomments,$WC,$WL);
23
24    }

```

No caso de o *post* não ter nenhum comentário ou nenhum *like* a relevância calculada é logo 0, pois não existiu nenhum utilizador a demonstrar interesse por aquele conteúdo partilhado. Caso contrário utilizando o método *statsByPost* é calculada a relevância de um dado conteúdo nesse grupo. O método *statsByPost* (5.4) calcula consoante o peso de cada componente a relevância desse *post*.

Listing 5.4: Método auxiliar para o cálculo da popularidade de um *post*

```

1
2
3   if ($meanlikes==0&&$meancomments!=0)
4   {
5       $r= (($comments/$meancomments) * $WC);
6   }
7   else if ($meancomments==0&&$meanlikes!=0)
8   {
9       $r= (($likes/$meanlikes) * $WL);
10  }
11  else if ($meancomments==0 && $meanlikes==0)
12  {
13      $r=0;
14  }
15  else{
16      $r= (($likes/$meanlikes) * $WL) + (($comments/$meancomments) * $WC);
17  }

```

A média de *likes* ou comentários tem de ser diferente de 0 para que se possa perceber qual o comportamento dos utilizadores no grupo senão a relevância será sempre nula. Pensemos que existe uma lista de posts sem que nenhum deles tenham *likes* ou comentários, nenhum desses *posts* terá relevância. Neste caso a variável W_l é 0.6 e a variável W_c é 0.4. Escolheu-se colocar estes pesos pois um *like* pode ser colocado apenas por um utilizador ou seja cada *like* é único num dado *post*. No caso do comentário podem existir vários comentários do mesmo conjunto de utilizadores e por isso decidiu-se dar ligeiramente mais valor a *likes*.

5.2.2.4 Cálculo da relevância do utilizador para o grupo

Para calcular a popularidade dos utilizadores num determinado grupo, é necessário executar o método *addStatUserGroup*. Basicamente este método vai somar as relevâncias de todos os *posts* e verificar de todos os utilizadores que partilharam quais é que tiveram uma maior relevância. No código presente na Listagem 5.5, demonstra-se como se procedeu à contagem dos *likes*, que é em tudo semelhante à dos comentários.

Listing 5.5: Contagem dos *likes* dos posts

```

1
2 foreach ($posts['userslikes'] as $l)
3 {
4     foreach($l as $s){
5
6         if(array_key_exists($s,$arraylikes)){
7
8             $arraylikes[$s]['likes']++;
9         }
10        else{

```

```

11         $arraylikes[$s]['likes']=1;
12     }
13
14 }
15 }

```

No caso das somas das relevâncias dos utilizadores, somaram-se todas as relevâncias de todos os utilizadores. No fim faz-se a razão entre a relevância de um utilizador e a relevância total.

5.2.2.5 Cálculo da relevância do grupo para o utilizador

Para calcular a relevância dos grupos para o utilizador, foi necessário que a relevância dos utilizadores estivesse calculada existindo assim um fluxo de dados seguido entre este processamento dos dados.

Inicialmente calculava-se há quanto tempo é que os membros de um grupo não partilhavam informação para os mesmos através do método *getDifferenceBetweenDates*. Depois dos dias calculados era usada uma expressão logarítmica para calcular o valor que efectivamente esta componente iria ter, de forma a que um o valor perca quantitativamente importância quanto há mais tempo foi feita a actualização. No código presente na Listagem 5.6 apresenta-se como foi implementado este processo.

Listing 5.6: Calculo do valor temporal

```

1
2 $days=0;
3 $date2 = new DateTime($date2);
4 $interval = $date1->diff($date2);
5
6
7 if($interval->m>=2 && $interval->d>=2)
8     return 0;
9 else {
10     if($interval->m==2)
11     {
12         $days=62;
13     }
14     if($interval->m==1)
15     {
16         $days=$days+31;
17     }
18
19     $days=$days+$interval->d;
20 }
21
22 if($days==0)
23 {
24     return 1;
25 }else

```

```

26 {
27     return log((62/$days))/log(62);
28 }
29 }

```

Depois calcula-se a relevância através das expressões onde é feito um cálculo ponderado de todas as variáveis.

Constantes utilizadas nas expressões na implementação

Existem várias constantes que foram utilizadas para o cálculo do peso de cada componente. Essas constantes representam o peso do interesse de um utilizador por um dado conteúdo partilhado. O valor que cada variável tem foi decidido através de vários testes computacionais, analisando os valores no próprio *Facebook* e os valores que as expressões davam.

Verificou-se a popularidade de um determinado *post*, verificando, se um *post* tinha bons indicadores de que era um conteúdo de interesse. Através da análise dos resultados obtidos com a aplicação e o que era esperado, conseguiu-se afinar os pesos usados nas expressões do cálculo da relevância. Os valores finais usados nas expressões são apresentados nas Tabelas 5.9 e 5.10.

Constante	Valor	Descrição
W_c	0.40	Valor da constante relativa ao peso que os comentários vão ter na expressão.
W_l	0.60	Valor da constante relativa ao peso que os <i>likes</i> vão ter na expressão.

Tabela 5.9: Constantes na expressão do cálculo da relevância de utilizadores

No caso da relevância de conteúdos, o valor W_l é ligeiramente superior ao W_c , pois cada *like* é só colocado por um utilizador enquanto diversos comentários podem ser colocados pelo mesmo utilizador. Dessa forma tem de ser valorizado o que consegue representar o maior número de utilizadores a demonstrar interesse.

Constante	Valor	Descrição
W_t	0.05	Valor da constante representativa do peso da actualização da informação no grupo;
W_l	0.25	Valor da constante representativa do peso dos <i>likes</i> colocados pelo utilizador;
W_c	0.25	Valor da constante representativa do peso dos comentários colocados pelo utilizador;
W_p	0.30	Valor da constante representativa do peso da relevância dos <i>posts</i> colocados pelo utilizador;
W_o	0.15	Valor da constante representativa do peso das relevâncias anteriores do utilizador.

Tabela 5.10: Constantes na expressão do cálculo da relevância de grupos para o utilizador

No caso da relevância dos grupos representado pela Tabela 5.10, não existe distinção entre as constantes representativas de *likes* e comentários, pois estes representam apenas

um utilizador em ambas as formas. A relevância dos *posts* que os utilizadores partilharam, tem um peso bastante acentuado, pois o facto de um utilizador partilhar conteúdos e os utilizadores demonstrarem interesse pela sua partilha, demonstra também que o utilizador tem interesse naquele grupo. Isto porque a ideia de que existem utilizadores que gostam do que se partilha num determinado grupo por norma também aumenta o interesse nesse grupo. O próprio facto de um utilizador partilhar, é ele próprio um factor de interesse pelo grupo.

Este peso para os *posts*, W_p , tornou-se o mais adequado, pois não se pretende dar um peso demasiado grande a esta constante, pois existem utilizadores que não partilham nada, mas são consumidores assíduos de informação. Em relação a W_o , peso da relevância que o utilizador teve no passado, 0.15 foi o valor que está abaixo do peso dos *likes*, *posts* e comentários que são informações com uma maior frescura e está acima do valor temporal. Quando se testou o aumento deste valor e conseqüentemente, a baixa dos restantes, o valor para que um utilizador tivesse relevância ficou bastante mais difícil de conseguir, pois estava demasiado peso nas relevâncias anteriores. Quando colocado este valor demasiado baixo, seria demasiado fácil obter uma relevância positiva pelo que este se afinou como o valor mais adequado. A constante temporal W_t , foi uma constante que teria a função de verificar que peso teria o facto de existirem conteúdos partilhados com frequência ou não. Como já foi referido na Secção 3.7.2 esta variável serve para dar peso aos grupos que efectivamente têm um nível de participação frequente pelo que o seu peso não poderia ser muito elevado. Dando demasiado peso a esta variável estava-se a tirar peso aos indicadores de interesse pelo que este valor 0.05 se tornou o mais adequado.

5.2.2.6 Sistema de *cache* do sistema

O cálculo dos dados que são obtidos através do *Facebook* só é realizado se essa informação já não estiver em memória. Para isso é necessário verificar quando foi a última actualização que ocorreu no grupo e verificar de quando é o último cálculo. O facto de termos esta implementação que funciona como uma *cache* permitiu que alguns problemas relacionados com os *timeouts* do servidor fossem resolvidos. Se não se implementasse esta forma estaria sempre a calcular para todos os utilizadores todas as relevâncias mesmo que os valores já tivessem sido calculados no passado e não houvesse actualizações à base de dados.

5.2.2.7 Controlador de extracção de tópicos

Este controlador é responsável por toda a parte da extracção de tópicos da informação que é obtida dos utilizadores e dos grupos.

5.2.2.8 Chamada ao servidor do *Alchemy*

A classe *Alchemy* é responsável por realizar as chamadas à API do *Alchemy* após receber os dados vindos do *Facebook*. Esta classe não implementa *Active record* pois não é necessário

guardar informação relativa às chamadas de uma à *API* na base de dados, sendo uma classe auxiliar.

Esta classe tem uma série de métodos implementados para proceder às chamadas ao *webservice* através do *cUrl* [cUR12]. Os métodos são todos semelhantes mudando os argumentos das chamadas. Por exemplo, no caso de ser uma entidade o argumento será *Text-GetRankedNamedEntities* enquanto de no caso de queremos uma categoria o argumento já será *TextGetCategory*. O exemplo de código de uma chamada ao *webservice* é o que se pode ver na Listagem 5.7.

Listing 5.7: Chamada ao servidor do *Alchemy*

```

1  $fields_string='';
2
3
4  $url = 'http://access.alchemyapi.com/calls/text/TextGetCategory';
5  $fields = array(
6      'apikey'=>urlencode('0d0ceb11d8c5635700b8e11e1113bd49c9187344'),
7      'text'=>urlencode($data),
8      'outputMode'=>urlencode('json')
9
10     );
11
12  //url-ify the data for the POST
13  foreach($fields as $key=>$value) { $fields_string .= $key.'='.$value.'&';}
14  rtrim($fields_string, '&');
15
16  //open connection
17  $ch = curl_init();
18
19  //set the url, number of POST vars, POST data
20  curl_setopt($ch,CURLOPT_URL,$url);
21  curl_setopt($ch,CURLOPT_POST,count($fields));
22  curl_setopt($ch,CURLOPT_POSTFIELDS,$fields_string);
23  curl_setopt($ch, CURLOPT_RETURNTRANSFER, TRUE);
24
25  //execute post
26  $result = curl_exec($ch);
27
28  //close connection
29  curl_close($ch);
30
31  return $result;

```

Quando se recebe os *posts* dos utilizadores ou quando se analisa o nome e as descrições dos grupos, tem de se ter em consideração, se existem dados suficientes para enviar a informação para o *Alchemy*. Isto é importante, pois o número de caracteres dos dados tem de ser superior a 15, para que o *Alchemy* consiga detectar qual a língua em que o *post* está escrito, senão coloca por omissão, o inglês [Alc12c]. Para analisar a informação vinda do *Facebook*, existem 6 diferentes métodos que permitem executar as chamadas às

APIs do servidor Alchemy. Esses métodos são:

- *do_post_requestEntities* - Analisa as entidades que podem ser encontradas nos textos dos *posts* dos utilizadores e também na informação relativa aos grupos;
- *do_post_requestRankedConcepts* - Analisa os conceitos que podem ser encontradas nos textos dos *posts* dos utilizadores e também na informação relativa aos grupos;
- *do_post_requestLink* - Analisa um determinado endereço que é recebido e tenta encontrar a categoria do mesmo conforme são encontrados na informação dos *posts* dos utilizadores e também na informação relativa aos grupos.
- *do_post_requestRankedWords* - Analisa as palavras chave que podem ser encontradas nos textos dos *posts* dos utilizadores e também na informação relativa aos grupos;
- *do_post_requestRelations* - Analisa as relações com certas entidades que podem existir com o texto recolhido dos *posts* dos utilizadores e também na informação relativa aos grupos;
- *do_post_requestCategory* - Analisa a categoria que pode ser encontrada nos textos dos *posts* dos utilizadores e também na informação relativa aos grupos;

No caso do método *do_post_requestLink*, optou-se apenas por tentar encontrar a categoria pois senão haveria demasiadas chamadas o que poderia baixar o desempenho da aplicação. O motivo pelo qual foi necessário implementar este método que analisa os endereços, é porque existiam apenas algumas correspondências entre grupos e *posts*. Grande parte das partilhas de informação no *Facebook* são partilhas de endereços externos ao *Facebook*, pelo que esta implementação foi mesmo necessária. No entanto mesmo apenas analisando a categoria, os resultados utilizando este método foram positivos como podemos ver na Secção 6.2.4.

5.2.2.9 Resposta do servidor do Alchemy

A resposta do servidor do *Alchemy* neste caso vem em formato JSON [JSOto]. Por exemplo no caso de um dado texto como, "No encontro de ontem com o presidente timorense, os prelados fizeram sentir a Matan Ruak que 'não é compreensível que nos nossos países exista tão elevado índice de pobreza'. No que toca ao caso de Timor-Leste - que o bispo de Baucau, D. Basílio do Nascimento, classifica como 'um país rico de gente pobre' -, o presidente solicitou à Igreja Católica 'ajuda para combater esse flagelo'." , o *Alchemy* devolve uma resposta em formato JSON sobre a categoria como o da listagem 5.8.

Listing 5.8: Resposta em formato JSON do servidor do *Alchemy*

```

1  "status": "OK",
2  "usage": "By_accessing_AlchemyAPI_or_using_information_generated_by_
    AlchemyAPI,_you_are_agreeing_to_be_bound_by_the_AlchemyAPI_Terms_of_Use
    :_http://www.alchemyapi.com/company/terms.html",

```

```
3     "url": "",
4     "language": "portuguese",
5     "category": "religion",
6     "score": "0.846572"
7 }
```

Uma resposta em *JSON* vindo da parte do *Alchemy*, tem campos criados que contêm informação para controlo como o *status*, que neste caso representa o sucesso da chamada ao servidor. Quando um método encontra por exemplo uma categoria, envia o valor que representa o grau de certeza de que a categoria está correcta. Nesta abordagem como as correspondências entre grupos e temas eram demasiado pequenas, não houve uma restrição a estes valores.

5.3 Camada de aplicação

Esta camada é composta por vários componentes como os *scripts*, que são interpretados e que facilitam a forma como os dados são expostos na interface, ou os próprios *templates* que estruturam a interface a ser mostrada. A pasta com os *templates* guarda a forma como as páginas vão ser mostradas na aplicação ao utilizador. Os *templates* são compostos com base na habitual linguagem de anotação como *HTML* e por *CSS* [W3C12b], que é uma linguagem para definir um estilo de uma página. Para ajudar os *templates* a apresentar a interface da aplicação, existem vários *scripts* em *JavaScript*². Outras bibliotecas implementadas em *JavaScript* com licença de código livre foram utilizadas nesta aplicação. Estas bibliotecas são :

- **JQuery** [Jqu12] - É uma biblioteca código livre implementada em *JavaScript* que simplifica a interpretação pela linguagem *HTML*. Esta biblioteca tem várias funcionalidades que facilitam a actualização de dados por exemplo por *AJAX* [Aja12] que permite o envio de dados para os *templates* da aplicação;
- **Amcharts** [Emp12] -É uma biblioteca código livre implementada em *JavaScript*, no entanto sem uma licença comercial não permite a sua usabilidade sem o aparecimento de um logotipo da empresa que a implementou. É um *script* utilizado para a criação de gráficos sobre a relevância de grupos e sobre a relevância dos utilizadores na aplicação. Um exemplo desta funcionalidade da interface pode ser visto na Figura 5.5.
- **Fancybox** [Fan12] - Esta biblioteca com licença de código livre em *JavaScript* permite mostrar imagens, conteúdos e outros componentes através de um aspecto organizado. Esta biblioteca permite abrir janelas com conteúdo, que se sobrepõe à janela que está a ser mostrada. Esta biblioteca é utilizada na funcionalidade de partilha de informação para os grupos.

²JavaScript - É uma linguagem bastante utilizada do lado do cliente que vai ter a tarefa de controlar e facilitar a forma como é criada a interface.

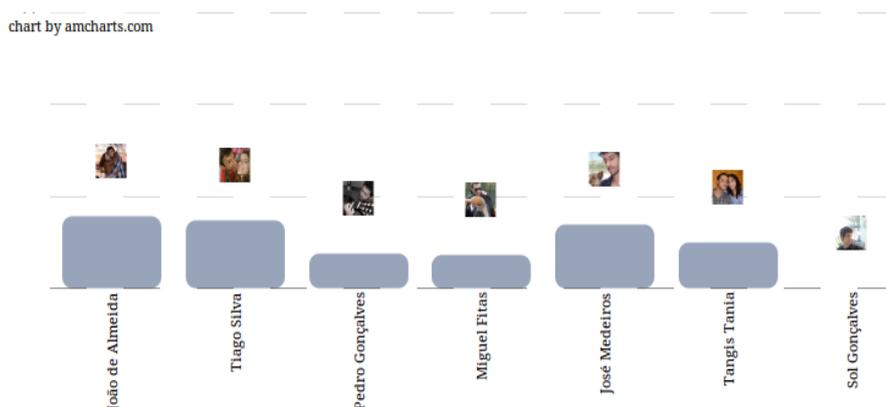


Figura 5.5: Gráficos da relevância de um utilizador no grupo

5.4 Camada de apresentação

Esta camada de apresentação é composta pela interface da aplicação no *Facebook*. É através desta camada de interface, que esta aplicação é mostrada aos utilizadores através de uma aplicação no *Facebook* [Fac12d]. Cada navegador (*browser*) de Internet mostrou diferentes comportamentos consoante as suas políticas de segurança. Neste caso existiram algumas dificuldades a nível da ligação segura. Se o utilizador de uma conta no *Facebook* utilizasse ligação segura para navegar no *Facebook*, alguns navegadores não permitiam que a aplicação GIA executasse pois inicialmente esta aplicação não tinha nenhum certificado de segurança, como foi o caso do *Firefox*. Ultrapassado os problemas de segurança, o *Facebook* envia para a aplicação GIA, a informação necessária sobre o utilizador e grupos, como por exemplo nome, imagem de perfil, que é tratada pela outras camadas mas mostradas na camada de interface. Após a instalação da aplicação GIA numa conta do *Facebook*, procede-se à execução da mesma, escolhendo a aplicação GIA da lista de aplicações.

Ao executar a aplicação irá aparecer uma janela de procedimento como se observa na Figura 5.6, enquanto se espera que toda computação de dados seja concluída, incluindo a execução das chamadas aos componentes externos.

Existem duas formas distintas de aceder à aplicação que são:

- Através do endereço da aplicação no *Facebook* - <https://apps.facebook.com/giatangis/>;
- Através de um endereço externo ao *Facebook* - <http://www.sporttools.org/tania/Gia/>.

Quando o processamento acaba, aparece uma janela com a informação no topo sobre o utilizador, onde aparece o nome, uma miniatura da imagem de perfil e a última vez que o estado do utilizador foi actualizado. Esta informação foi colocada de forma redundante, para o caso de não se aceder a partir do perfil do utilizador, ou seja no caso de um utilizador querer aceder a esta aplicação apenas utilizando um endereço, conseguir visualizar que utilizador do *Facebook* está a ser considerado para executar a aplicação.



Figura 5.6: Imagem do processamento da aplicação *GIA*

Na interface da aplicação aparece também uma lista de opções (*menu*) com as diversas páginas onde se mostrará os resultados ao utilizador. Nas secções seguintes se descreverá cada funcionalidade da interface proposta.

5.4.1 Estatísticas (*Stats*)

Na opção *Stats*, podemos encontrar informação sobre as estatísticas dos grupos do utilizador que são executados através da aplicação *GIA*. Estas estatísticas mostram várias informações sobre os grupos como a média de *likes* ou comentários existentes num determinado grupo ou o número de utilizadores diferentes que partilham informação. A visualização desta opção pode ser vista na Figura 5.7.

5.4.2 Relevância dos grupos (*Groups relevance*)

Na opção *Groups relevance* podemos encontrar dados sobre a importância dos grupos para o utilizador. Existe uma lista de grupos, com os indicadores de importância que um utilizador colocou naqueles grupos ou seja se o utilizador, mesmo não partilhando um *post* no grupo, pode ter colocado um *like* ou comentário e isso é contabilizado. O resultado se um grupo é efectivamente importante ou não para o utilizador é mostrado através de uma mensagem e através dos *ícones* do *Facebook* de *gosto* ou de *não gosto* (*like/dislike*) como podemos ver na Figura 5.8.

Existe também um gráfico com as relevâncias para o utilizador de cada grupo onde podemos comparar de uma forma mais simples os diversos grupos. Os grupos cuja relevância é nula, são excluídos nesse gráfico. Através da Tabela 5.11 mostramos os níveis de relevância que foram considerados.

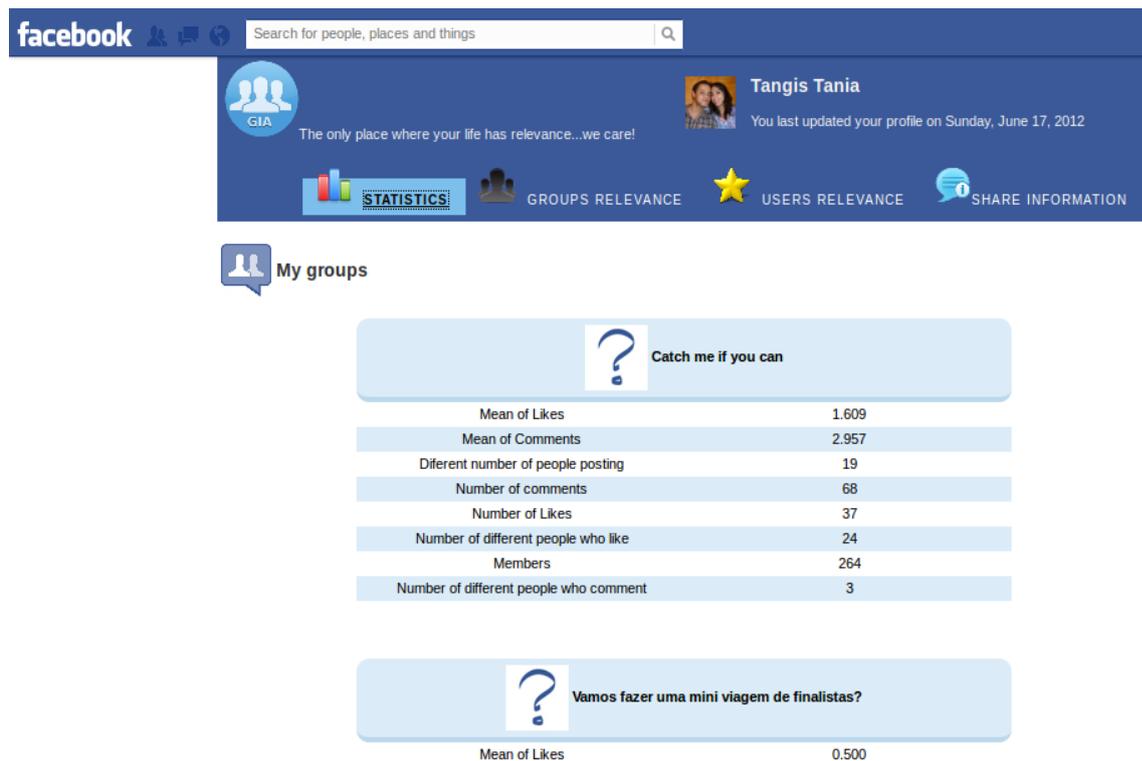


Figura 5.7: Janela da aplicação GIA com as estatísticas dos grupos

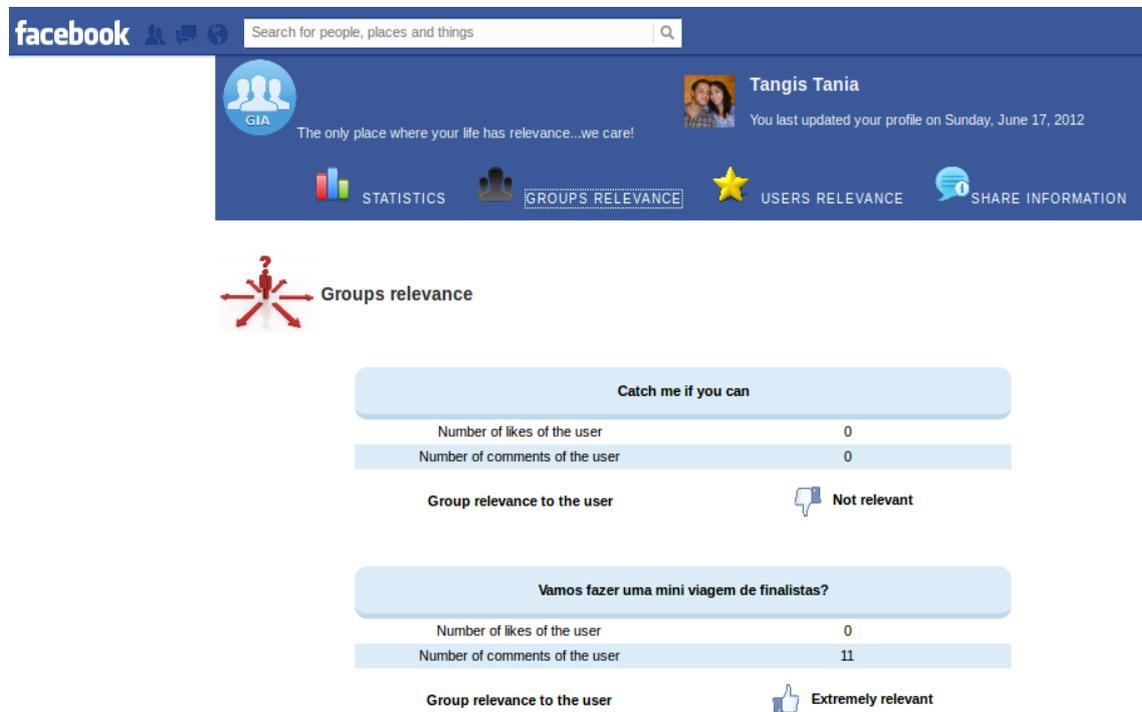


Figura 5.8: Janela da aplicação GIA com a relevância dos grupos para um utilizador

Legenda	Intervalo de valores
Nada relevante	<0.01
Pouco relevante	0.01-0.09
Relevante	0.09-0.20
Muito relevante	0.20-0.35
Extremamente relevante	>0.35

Tabela 5.11: Legenda para os valores da relevância dos grupos para o utilizador

Estes intervalos de valores foram atribuídos através de vários testes aos dados devolvidos pela aplicação e pela verificação de qual era efectivamente o grau de relevância que os grupos tinham.

5.4.3 Relevância dos utilizadores (Users relevance)

Na opção *Users relevance* aparecem tabelas e listas de gráficos com a importância dos utilizadores para o grupo. As tabelas têm os valores concretos e os gráficos excluem os utilizadores com relevância nula para se perceber mais facilmente quem são os utilizadores que efectivamente são importantes. Esta opção também ajuda a que o gráfico não fique demasiado grande. Podemos ver um exemplo desta página na Figura 5.9.

5.4.4 Partilha de informação (Share Information)

Na opção *Share information* aparece uma janela com uma lista de *posts* dos utilizadores e uma lista de grupos. A aplicação sugere, através da correspondência entre temas dos grupos e dos conteúdos dos *posts*, quais devem ser partilhados e para que grupos.

No caso de existir informação para o *matching*, os *posts* são mostrados aos utilizadores através de uma janela (*Dialog*), que aprova os *posts* para os grupos.

Quando carregamos no botão de partilha (share) os *posts* que forem escolhidos através da *checkbox* são publicados dentro dos grupos escolhidos pelo utilizador. Como exemplo desta janela temos a Figura 5.10.

5.5 Resumo

Após ultrapassadas todas as dificuldades da programação e de um exaustivo período de testes foi possível terminar a implementação desta aplicação. Através do uso desta aplicação foi possível obter sugestões sobre o que poderia surpreender os utilizadores e perceber o que realmente eles gostariam de ter numa aplicação deste tipo. Os resultados obtidos, a sua análise e as sugestões dadas por utilizadores desta aplicação, são descritos com mais pormenor no Capítulo 6, onde se apresentam os dados experimentais e os questionários.

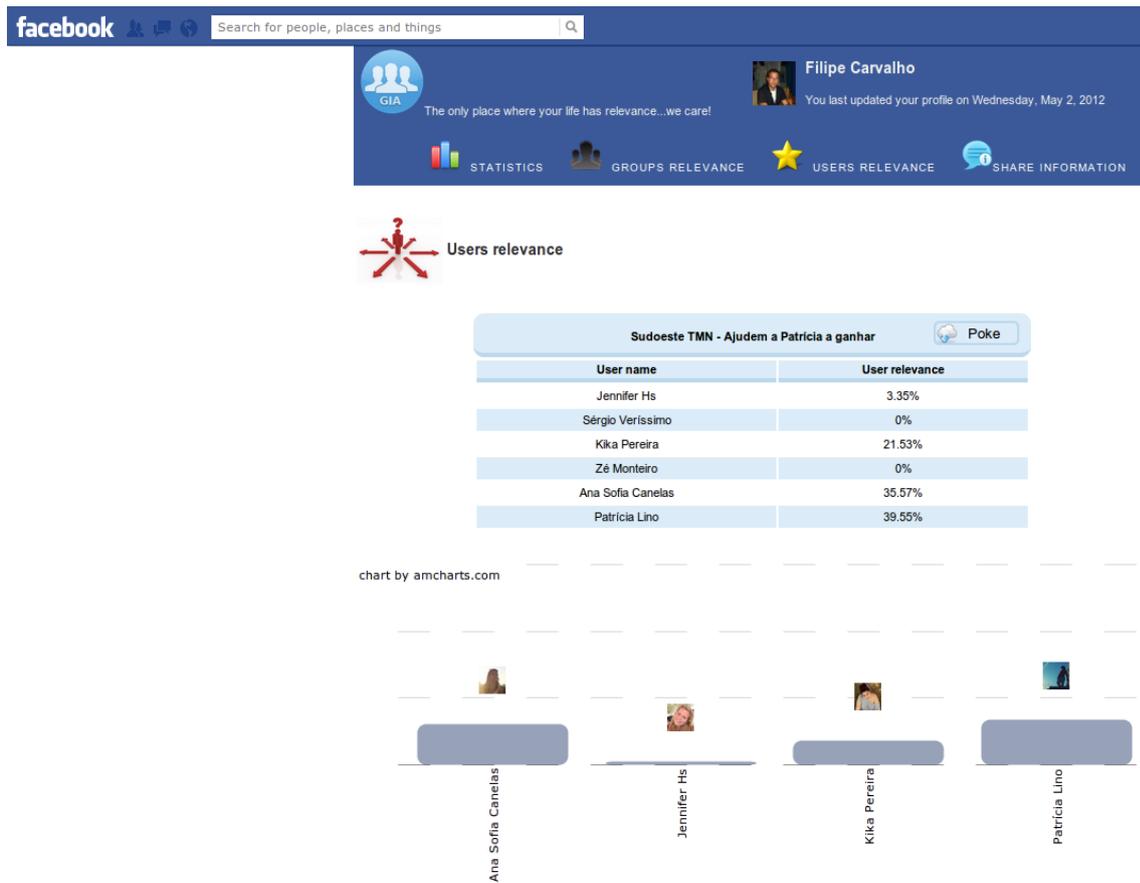


Figura 5.9: Janela da aplicação GIA com as importâncias dos utilizadores nos grupos

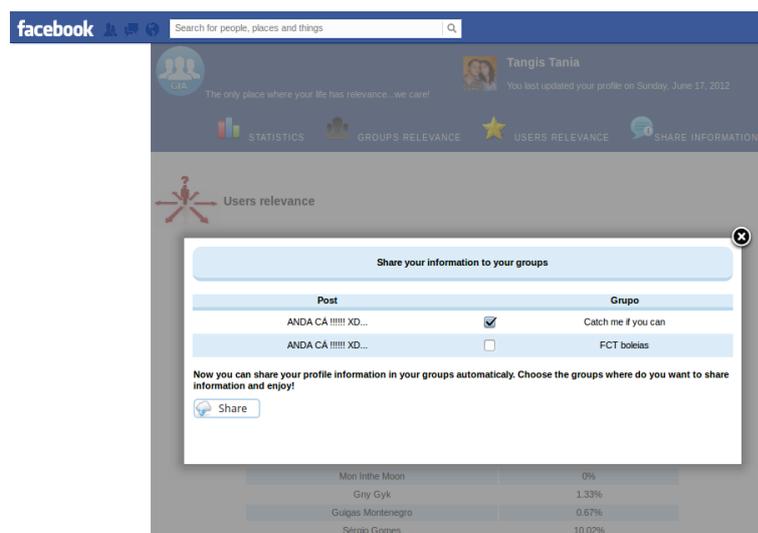


Figura 5.10: Janela da aplicação GIA com a partilha de informação nos grupos

6

Avaliação experimental

Através do desenvolvimento do protótipo da aplicação *GIA*, foi possível obter dados sobre os utilizadores e os grupos, assim como a opinião de vários inquiridos sobre a sua experiência com a aplicação. Registaram-se algumas sugestões e opiniões sobre os dados mostrados aos utilizadores pela aplicação, de forma a se poder adaptar a aplicação a valores que estejam mais correctos, modelar as expressões propostas no Capítulo 3, mas também para se conseguir perceber o que os utilizadores esperavam da aplicação. Nas próximas secções descrevem-se quais os resultados obtidos pela aplicação, a análise de dados, os inquéritos de satisfação e as conclusões sobre toda a informação extraída.

6.1 Descrição do ambiente e universo de utilizadores

Durante e após a fase de implementação foram obtidos e analisados resultados durante cerca de 1 mês e meio através de 126 utilizadores distintos.

Através das estatísticas do *Facebook* e da aplicação *GIA*, guardadas na base de dados da aplicação, foi possível perceber que tipo de utilizadores utilizaram a aplicação, podendo extrair informação estatística, como por exemplo, quais os escalões etários ou o género de quem utilizou a aplicação. A Figura 6.1 mostra as estatísticas da aplicação *GIA*, que o *Facebook* disponibiliza para qualquer aplicação. Como se pode observar esta aplicação foi utilizada maioritariamente por utilizadores que habitam em Portugal, mas existiram 12 utilizadores de outros países que utilizaram a aplicação. Cerca de 72% dos utilizadores são do sexo masculino enquanto que 28% são do sexo feminino. A publicação dos resultados obtidos a partir da aplicação obteve o consentimento de todos os utilizadores, que instalaram a aplicação e aceitaram a autorização da sua instalação

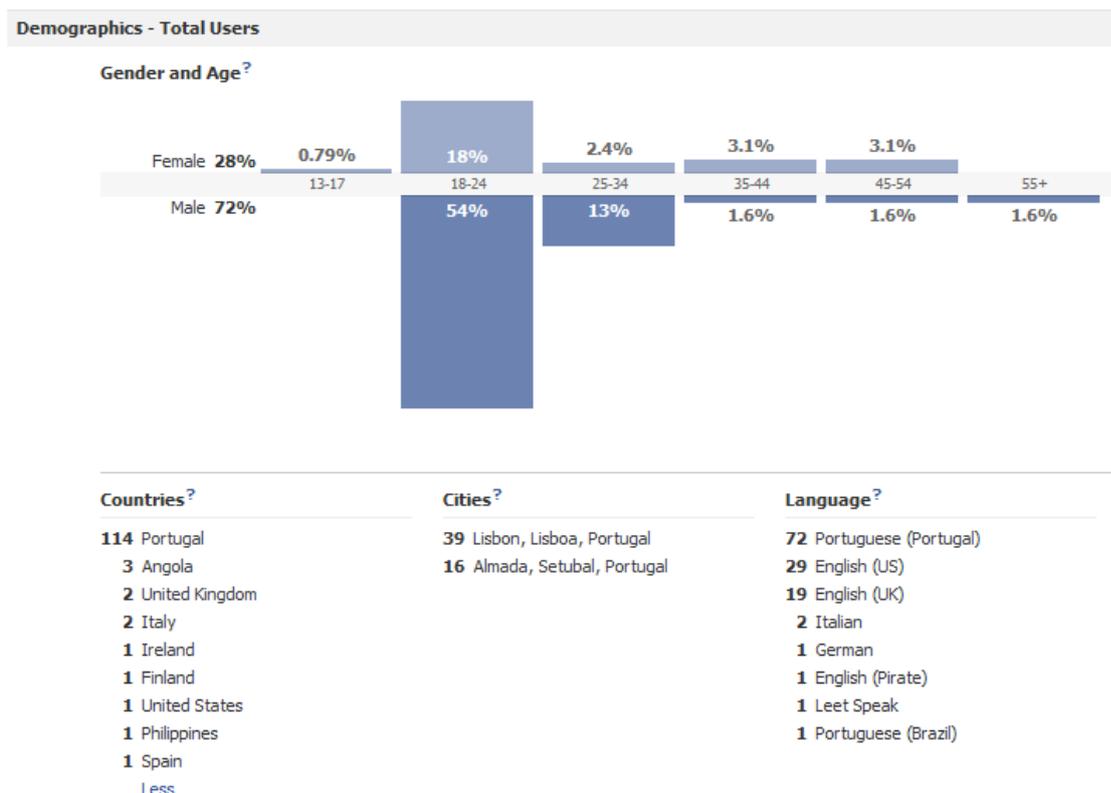


Figura 6.1: Estatísticas do *Facebook* da aplicação *GIA*

através do *Facebook*. Esta aplicação foi executada por vários utilizadores cerca de 876 vezes, durante 3 semanas desde a terceira semana de Julho à primeira de Agosto, depois da fase de testes ter sido concluída. Um conjunto de utilizadores mais próximos, permitiram também, numa fase inicial, fazer um extensivo processo de testes (*debugging*) à aplicação tendo sido resolvidos todos os problemas encontrados, graças às opiniões dadas pelos utilizadores. Através dos utilizadores foi possível obter informação sobre 723 grupos distintos. Relativamente ao número de grupos de cada utilizador verificou-se que existia uma grande diversidade observando-se que existem utilizadores que apenas pertenciam a 2 ou 3 grupos, enquanto que outros pertenciam a 30 ou mais. Segundo Mislove [MMG⁺07], isso deve-se ao facto de os utilizadores mais socializáveis, por norma terem um maior número de grupos. Dai que embora existam 126 utilizadores, existem muitos mais grupos analisados. Os grupos analisados foram divididos em 3 classes consoante o número de utilizadores, como mostra a Figura 6.2, considerando que um grupo pequeno está entre 1 e 149 membros, um grupo médio entre 150 e 599 utilizadores e um grupo grande com mais de 600 utilizadores.

No âmbito destes grupos verificou-se a relevância de 5472 utilizadores que participaram na partilha de informação no contexto dos grupos e abrangeram-se 516291 utilizadores pertencentes aos grupos.

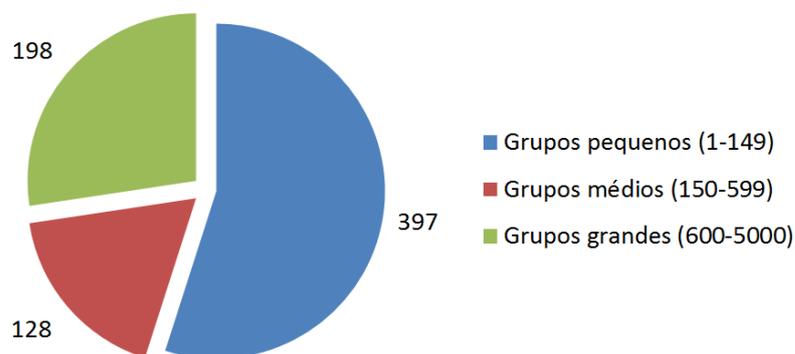


Figura 6.2: Distribuição de grupos analisados no GIA

6.2 Resultados experimentais

Nesta secção apresentam-se alguns dos resultados obtidos através da aplicação e uma análise de forma a identificarmos possíveis padrões de comportamento por parte dos utilizadores nos grupos. Os testes a utilizadores começaram na terceira semana de Julho tendo -se intensificado na última semana de Julho e terminaram na segunda semana de Agosto, para os efeitos deste documento.

Analisaram-se vários comportamentos dos utilizadores nos vários tipos de grupos. Analisando a população de utilizadores que partilham informação para os grupos, vão-se analisar quais os rácios de participação habituais e quais os utilizadores, que ao participar não tiveram uma participação relevante. Para isso apresentar-se-á um conjunto de exemplos mais específicos e também mais genéricos para comentar os valores obtidos. Também haverá exemplos sobre a popularidade de utilizadores nos grupos e a importância que os grupos tiveram para dois utilizadores, analisando a quantidade de grupos a que os utilizadores não dão relevância alguma.

6.2.1 Comportamento dos utilizadores nos grupos

Como já foi referido na análise do comportamento dos utilizadores dentro dos grupos achamos interessante identificar 3 tipos de grupos, nos quais os comportamentos dos utilizadores são distintos.

As Tabelas 6.1 e 6.2, mostram o número de *posts* extraídos no grupo e a diversidade de utilizadores que os partilharam numa execução da aplicação GIA. Existem vários tipos de grupos nestas tabelas representados, criados devido a diferentes âmbitos. No entanto, este estudo tem por objectivo focar em grupos que tem a sua actividade regular e não em grupos que apenas têm a sua actividade na altura perto de um evento como acontece com a "Festa do Avante" por exemplo. No caso do número de participantes na parti-

Grupo	Membros	Número de <i>Posts</i>	Número de participantes a partilhar
Andanças	4842	25	16
<i>Just Been Paid</i>	894	20	7
Produtora de Eventos- Brand	1799	25	2
Social Erasmus - Lisboa	180	25	9
Mastodon Portugal	243	25	11
Escola secundária João de Barros	409	20	13
Concurso de Programação FCT	21	24	9
aNTUNIA	74	18	12
IRC-LEI	101	24	10
Boleia e Afins	11	17	8

Tabela 6.1: Participação nos diferentes tipos de grupos no *Facebook*

lha de conteúdos em relação ao número de membros, existem mais utilizadores distintos a participar nessa actividade, em grupos mais pequenos, do que em grupos médios ou grandes. O número de *posts*, representa os *posts* dos grupos que se puderam obter, através da *API* do *Facebook*. No entanto, existem grupos em que os utilizadores nunca partilharam mais de 25 *posts* (número limite definido através da *API* do *Facebook*), motivo pelo qual aparecem números inferiores a 25 *posts*. Nesses casos é comum, os grupos com menos de 25 *posts* publicados, serem grupos que a sua actividade é limitada devido à sua criação ter um objectivo limitado de tempo.

Grupo	Utilizadores a comentar	Utilizadores a gostar	Rácio de gostos	Rácio de comentários
Andanças	5	24	0.0076	0.0080
<i>Just Been Paid</i>	3	8	0.0167	0.0257
Produtora de Eventos- Brand	1	18	0.027	0.00167
Social Erasmus - Lisboa	4	20	0.144	0.138
Mastodon Portugal	2	14	0.168	0.0452
Escola secundária João de Barros	2	14	0.0415	0.0293
Concurso de Programação FCT	8	10	1	7.14
aNTUNIA	6	20	0.5	2.054
IRC-LEI	2	13	0.237	0.069
Boleia e Afins	5	4	0.818	6.54

Tabela 6.2: Participação nos diferentes tipos de grupos no *Facebook* II

Verifica-se que os grupos mais pequenos têm uma maior actividade dos seus membros do que os grupos maiores. Muitas vezes isso se deve ao facto de as pessoas serem mais cuidadosas a partilhar informação para muitos utilizadores do que para um grupo mais restrito de utilizadores. Nestes grupos foi possível também verificar que existe sempre um pequeno grupo de participantes que é bastante activo e que a grande parte dos membros é apenas passiva ou seja assumimos que apenas se limitam a consultar a informação partilhada. Quanto maior for o número de membros de um grupo mais este comportamento se verifica. O rácio entre o número de *likes* e o número de comentários vai variando com a dimensão do grupo, também verificando o mesmo comportamento.

O rácio vai aumentando na medida em que, quanto mais pequeno for o grupo, mais utilizadores participam. Em grupos mais pequenos, a proporção de utilizadores a gostar ou a comentar é maior que em médios ou grandes. Por exemplo, na Tabela 6.1 podemos observar que, apenas 16 utilizadores do grupo do festival Andanças participam na actividade de partilha, de entre os 4842 membros. Isto poderá ter haver com o objectivo pelo qual este grupo foi criado. Ainda como exemplo, no grupo mais pequeno, "Boleia e afins", 8 dos 11 membros do grupo partilham informação. Em relação à popularidade dos utilizadores e ao tipo de participação no contexto de grupos, existe um comportamento ligeiramente diferente nas várias classes de grupo, como se discute nas secções seguintes.

6.2.2 Relevância de utilizadores para os grupos

Durante o tempo de testes foram armazenadas informações relativas às relevâncias e às participações de cada utilizador nos grupos. As informações armazenadas na base dados são dados estatísticos sobre os *posts*, os seus criadores e qual a relevância que esse *post* teve no âmbito de um determinado grupo. Essa relevância de *posts* foi calculada através das expressões do Capítulo 3 e 4. Através destes dados, é possível verificar neste intervalo de tempo, quantos utilizadores partilharam informações para o grupo e quais é que não tiveram qualquer popularidade.

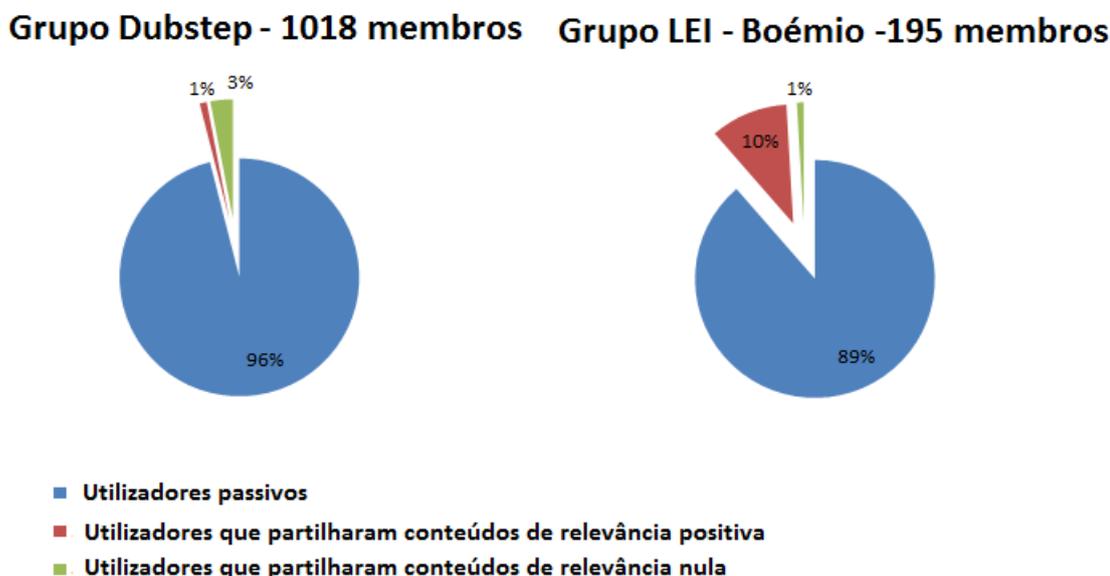


Figura 6.3: Participação de utilizadores em vários grupos específicos I

Nas Figuras 6.3 e 6.4 temos o exemplo de 3 grupos de classes diferentes e dos níveis de participação encontrados.

Para não se tirarem conclusões baseadas em casos eventualmente muito específicos, contou-se o número de todos os membros de grupos em todas as classes, ou seja, pequenas, médias e grandes. Quantificou-se também o número de utilizadores que partilharam

Grupo Mixórdias de Informática - 46 membros

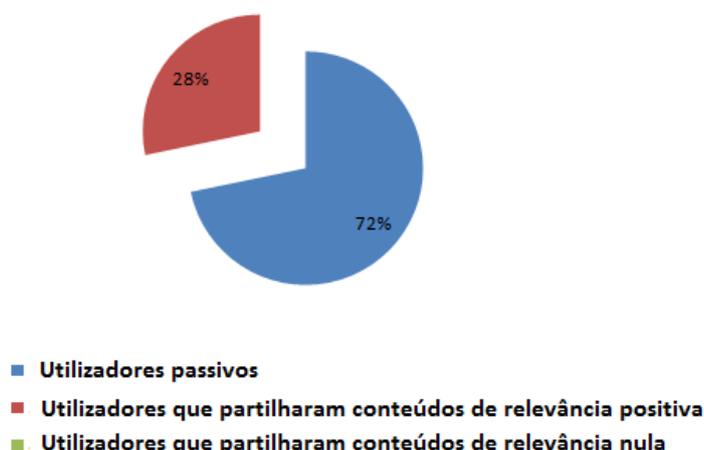


Figura 6.4: Participação de utilizadores em vários grupos específicos II

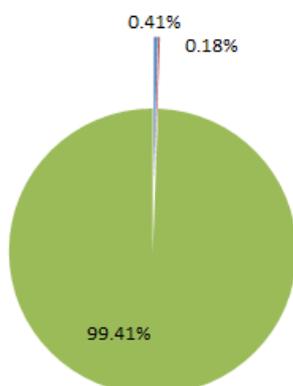
informação para os grupos e quais desses utilizadores tiveram relevância nula. Os resultados são mostrados nas Figuras 6.5 e 6.6 e foram semelhantes, confirmando a tendência de nos grupos maiores haver um maior número de utilizadores passivos e menos utilizadores a partilhar.

O número de utilizadores que foram contabilizados pela aplicação como não relevantes é maior quando se aumenta o número de utilizadores. Um utilizador é considerado não relevante quando partilha conteúdos num grupo, em que nenhum membro do grupo demonstrou interesse, tal como foi referido no Capítulo 3. O número de utilizadores não participantes também vai aumentando nesse caso. Mostra-se assim que existem utilizadores que têm mais relevância que outros. Particularmente num grupo médio e grande existe uma significativa diferença entre os mais populares e os menos populares. Apesar de um utilizador num grupo maior ter mais dificuldade em ter uma grande relevância por existirem mais utilizadores, o número de utilizadores com uma grande popularidade é reduzido.

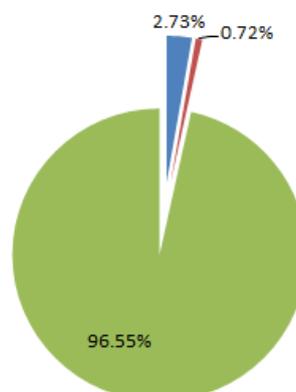
Existem algumas diferenças entre o caso mais específico (Figuras 6.3 e 6.4) e o geral (Figuras 6.5 e 6.6), que mostra que no caso específico houve mais pessoas a partilhar sem obter qualquer *feedback*, o que poderá ter em conta o âmbito desse grupo. Neste caso geral a mesma tendência de actividade se encontra em ambos os casos e em ambas as dimensões de grupo, em que os grupos mais pequenos têm mais actividade e mais utilizadores a partilharem conteúdos.

Em relação à popularidade de um utilizador para o grupo, o facto de permitir que um utilizador tenha acesso à sua popularidade, provoca uma tendência de maior número de utilizadores a quererem participar num determinado grupo.

Participação de utilizadores em grupos de grandes dimensões



Participação de utilizadores em grupos de médias dimensões



- Utilizadores que partilharam conteúdos de relevância positiva
- Utilizadores que partilharam conteúdos de relevância nula
- Utilizadores passivos

Figura 6.5: Participação em vários tipos de grupos I

Uma consequência encontrada é que o facto de os utilizadores verem as suas popularidades, motiva a participação e começa a existir uma competição entre utilizadores para obterem uma maior popularidade que outros, partilhando informação com conteúdos considerados mais relevantes e que interessem às propriedades do grupo. Esta competitividade leva a que o grupo tenha informação com mais qualidade devido à grande selectividade dos conteúdos que os utilizadores partilham como em [TR10]. Por exemplo, um utilizador vai seleccionar melhor a informação antes de a partilhar para o grupo, para que a sua reputação não diminua. Este comportamento observou-se através de um grupo específico de utilizadores de testes mais próximos.

O facto de um grupo ser pequeno e ter os objectivos do mesmo bastante definidos, aumenta a probabilidade de um determinado conteúdo partilhado agradar aos membros dos grupos. Quando existem muitos utilizadores, torna-se mais difícil que um conteúdo agrade a vários utilizadores. Na Figura 6.7 está um exemplo do grupo INOV com 4951 membros e da popularidade dos seus participantes. A relevância dos participantes neste grupo está mais ou menos ao mesmo nível, excepto 3 utilizadores que se destacam com uma maior popularidade.

Ainda como exemplo temos dois grupos mais pequenos para se perceber como normalmente é distribuída a popularidade nesses grupos nas Figuras 6.8 e 6.9. O grupo de *Jantar de Natal de 2011* tem 9 membros enquanto que o grupo *Tasqueiros de Dresden* têm 31 membros. No grupo *Tasqueiros de Dresden* existe uma uniformidade na popularidade dos utilizadores que partilham conteúdos não existindo nenhum que se destaque muito significativamente. Isto mostra que existem vários comportamentos nos grupos, dependem do objectivo pelo qual o grupo foi criado. No entanto o mais habitual é existir um número

Participação de utilizadores em grupos de pequenas dimensões

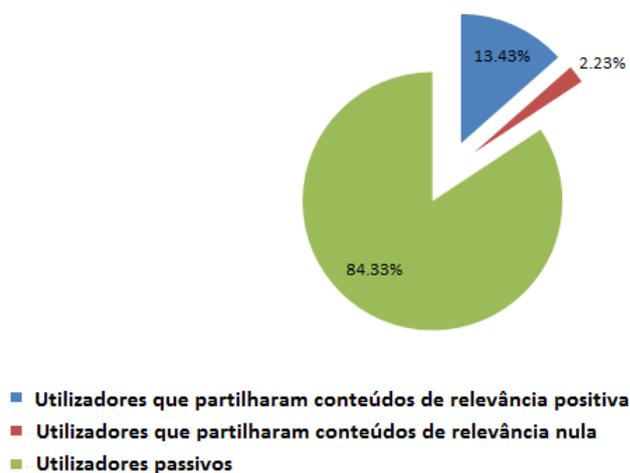


Figura 6.6: Participação em vários tipos de grupos II

muito reduzido de utilizadores (cerca de 2 ou 3) que se destacam mais como no caso do grupo *Jantar de Natal de 2011*, grupo criado com objectivo de dar apoio à realização de um jantar de Natal.

6.2.3 Importância dos grupos para os utilizadores

Uma outra dimensão interessante desta aplicação passa pela avaliação da relevância dos grupos para os próprios utilizadores. Nos ensaios realizados, a aplicação *GIA* calculou e guardou os grupos analisados assim como a sua popularidade para os utilizadores, o que irá permitir avaliar a quantidade de grupos que são importantes e os que não são. A título indicativo, dos 718 grupos em que foi possível avaliar a relevância dos grupos para os utilizadores, verificámos que existem 239 grupos cuja relevância se situa entre 0 a 3%. Ou seja cerca de 33% dos grupos existentes nesta análise, poderiam ser excluídos pois os seus utilizadores não partilham, gostam ou comentam conteúdos. É um número bastante elevado, o que leva a pensar na quantidade de grupos inactivos que um utilizador tem associado à sua conta de *Facebook*. No entanto existe a consciência de que muitos grupos estão inactivos pois apenas são utilizados 1 vez por ano, por exemplo, no caso de grupos associados a eventos localizados em datas fixas, por exemplo, como festivais ou conferências anuais. No entanto durante esse espaço de tempo, esse grupo, ainda que praticamente inactivo, continua associado à conta do *Facebook* do utilizador.

Avaliando os utilizadores de teste podemos verificar que existem diferentes padrões de utilizadores, mais activos e mais passivos. Foram escolhidos do conjunto de utilizadores da aplicação *GIA*, 2 utilizadores com diferentes comportamentos, sendo que um tem um comportamento activo, com várias participações dentro dos grupos e um mais

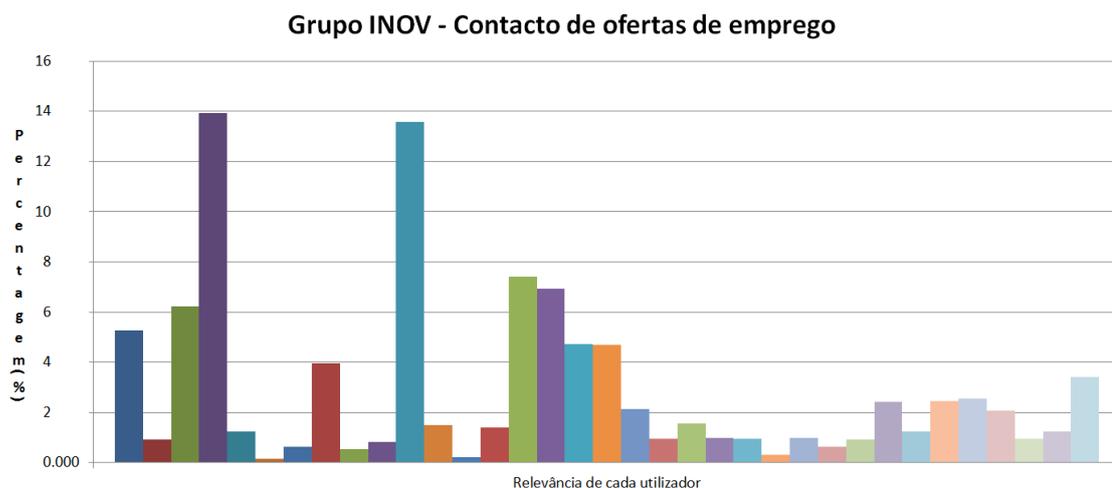


Figura 6.7: Grupo *INOV* analisado através do *GIA*

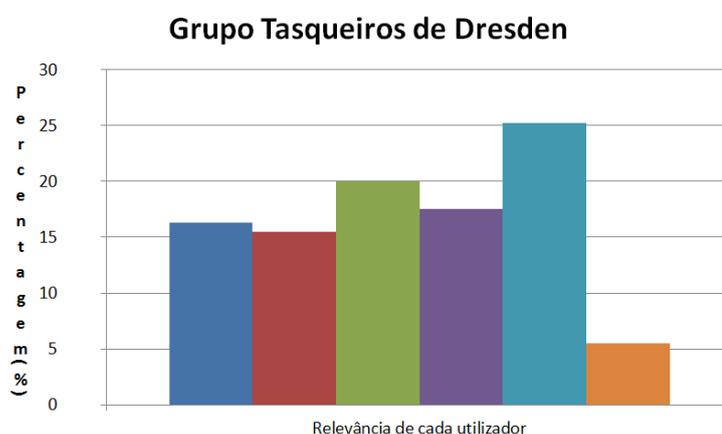


Figura 6.8: Grupo *Tasqueiros de Dresden* analisado através do *GIA*

passivo sem praticamente participações nos grupos. O número de grupos em que a importância dos grupos para o utilizadores é reduzida, é bastante maior quando os utilizadores são mais passivos, pois muitas vezes entram em grupos, mas não consultam nem partilham informação para o mesmo, sendo como utilizadores fantasma. As Figuras 6.10 e 6.11 mostram os gráficos da relevância dos grupos pessoais de 2 utilizadores distintos.

O utilizador João¹ é bastante mais activo nos grupos que a utilizadora Maria¹. Existe um conjunto de grupos com a percentagem de relevância bastante semelhante, pois são grupos activos em que o utilizador apenas se mostra passivo e consumidor de conteúdos, sendo que a percentagem de relevância obtida, praticamente se deve ao facto de esse grupo ser activo ou seja existirem outros utilizadores a partilhar informação.

¹Apesar dos utilizadores terem permitido extrair os dados em relação aos grupos através da aceitação desta aplicação, não se identifica o nome real dos utilizadores para o qual a relevância dos grupos foi calculada.

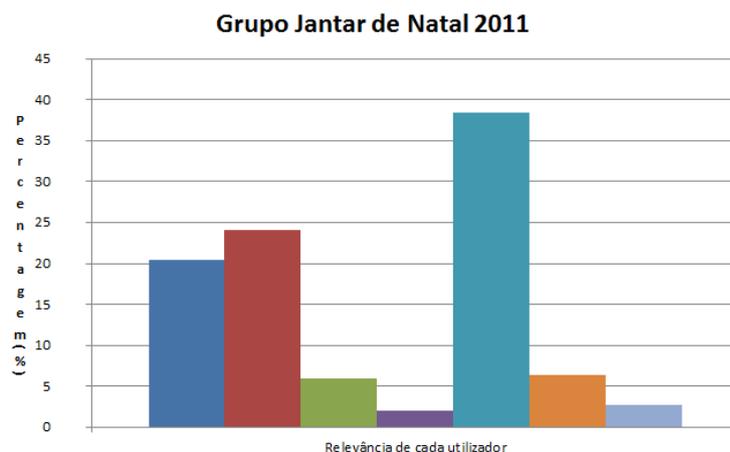


Figura 6.9: Grupo *Jantar de Natal 2011* analisado através do GIA



Figura 6.10: Relevância de grupos para o utilizador João

A nova funcionalidade do *Facebook* "*Seen by*" [Fac12h] poderá ajudar a melhorar estes resultados, em relação à identificação da consulta de informação. Esta funcionalidade do *Facebook*, permite contabilizar e identificar que utilizadores consultaram os *posts* partilhados nos grupos. No entanto essa funcionalidade não é acessível através da *API* do *Facebook*.

6.2.4 Correspondência entre conteúdos e grupos

Em relação à correspondência entre grupos e os conteúdos dos *posts* nos grupos, existem alguns dados interessantes.

Os dados foram obtidos através da análise das informações armazenadas na base de dados, verificando, quais os grupos que tinham informação suficientes para serem analisados e serem usados para fazer correspondência. No caso de não existir informação suficiente, por exemplo não existindo uma descrição do grupo, nenhuma informação sobre as categorias e as outras características do grupo encontradas pelo *Alchemy* eram armazenadas na base de dados.

Se a informação extraída dos grupos do *Facebook*, for demasiado ambígua, o *Alchemy*

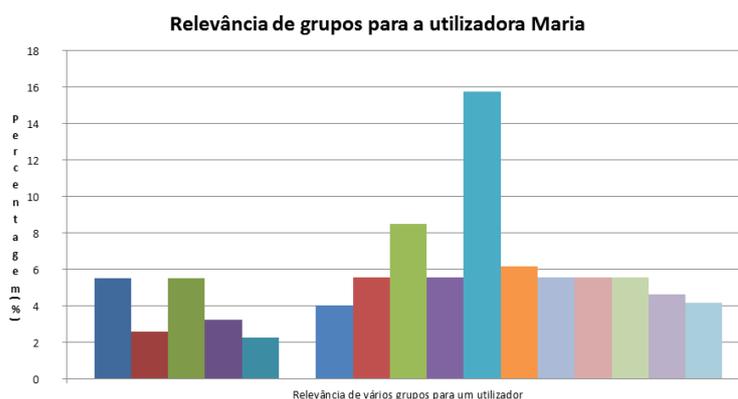


Figura 6.11: Relevância de grupos para a utilizadora Maria

não consegue extrair qualquer tópico, pelo que não vai existir informação na base de dados sobre esse grupo a não ser o nome, membros e a data da última actualização. Baseado nos dados acima referidos, dos 723 grupos analisados, apenas 223 têm informação suficiente para serem analisados, ou seja, para que se consiga perceber para que fim aquele grupo foi criado através das propriedades do perfil de grupo. Grande parte destes grupos, apenas têm nome e não têm descrição. Quando isso acontece esses grupos são à partida excluídos para não estarem a roubar tempo de computação ao servidor. O facto dos grupos não terem descrições, pode dever-se ao facto do *Facebook* apenas permitir colocar a descrição uma vez, sem que seja possível mudá-la futuramente. Destes 223 grupos que tinham informação suficiente, apenas em 112 foi possível descobrir qual era efectivamente o tema do grupo, através do *Alchemy*, pois a descrição não era suficientemente elucidativa e não era possível extrair tópicos e temas através da descrição. Alguns grupos têm na descrição o próprio nome, por exemplo, face a um grupo chamado “Vamos arriscar”, se a descrição for igual ao nome, fica-se sem se perceber qual o âmbito do grupo.

Em relação à própria partilha de informação para os grupos, os utilizadores decidiram partilhar pelo menos 62 vezes *posts* para os grupos. Existiram 473 correspondências entre *posts* e grupos o que é um valor ainda assim elevado. Esta funcionalidade é merecedora de uma maior análise no futuro, com vista a melhorar os resultados obtidos.

6.3 Inquéritos de satisfação e análise de resultados

Após a conclusão dos testes e com a aplicação completamente operacional, pediu-se aos utilizadores que demonstrassem o seu grau de satisfação através de inquéritos. Os inquéritos de satisfação começaram a realizar-se no dia 20 de Junho e estão publicados na própria aplicação, pelo que não houve necessidade de acompanhamento no preenchimento do questionário. Do total de 126 utilizadores que instalaram a aplicação, apenas 48 responderam ao inquérito.

6.3.1 Inquérito de satisfação

Os inquéritos de satisfação realizados aos utilizadores que utilizaram a aplicação, tinha como principal objectivo não só a acreditação das expressões sobre a relevância da informação, utilizadores e grupos, apresentadas no Capítulo 3, mas também averiguar que importância davam a uma ferramenta que tivesse estas funcionalidades. O inquérito foi criado através do serviço de inquéritos da empresa *Google* [Goo12a] e era composto por 15 perguntas sobre a implementação das expressões e ideias do Capítulo 3 mas também em relação à própria aplicação e aos resultados demonstrados. As perguntas realizadas foram as seguintes:

1. Utiliza *sites* de redes sociais? (Nunca, raramente, semanalmente, mensalmente, diariamente);
2. Que *sites* de redes sociais usa? (Resposta aberta);
3. Considera importante o uso de uma ferramenta para indicar a relevância dos utilizadores para os grupos? (Resposta na escala de 1 a 5 em que 1 é valor mais baixo e 5 valor mais alto);
4. Porque é que considera importante a ferramenta mencionada na questão 3. (Resposta aberta);
5. Considera importante uma ferramenta que detecte quais os grupos que já não têm interesse para si? (Resposta na escala de 1 a 5 em que 1 é valor mais baixo e 5 valor mais alto);
6. Em relação ao *GIA*, os resultados da importância dos utilizadores para o grupo correspondem ao que esperava? (Resposta na escala de 1 a 5 em que 1 é valor mais baixo e 5 valor mais alto);
7. Ainda em relação ao *GIA*, os resultados obtidos na relevância dos grupos para o utilizador correspondem ao que se esperava? (Resposta na escala de 1 a 5 em que 1 é valor mais baixo e 5 valor mais alto);
8. Considera importante uma funcionalidade que sugira a partilha da sua informação pública (*posts*) para os grupos com base no tema de cada grupo? (Resposta na escala de 1 a 5 em que 1 é valor mais baixo e 5 valor mais alto);
9. No caso do *GIA*, acha que as correspondências entre a sua informação e seus grupos reflectem o que desejaria partilhar? (Resposta na escala de 1 a 5 em que 1 é valor mais baixo e 5 valor mais alto);
10. No geral achou fácil de utilizar a aplicação *GIA*? (Resposta na escala de 1 a 5 em que 1 é valor mais baixo e 5 valor mais alto);
11. Que novas funcionalidades sugeria que a ferramenta tivesse? (Resposta aberta);

12. Que outros dados sugeria que a ferramenta disponibilizasse? (Resposta aberta).

Para além destas respostas o utilizador respondeu também a perguntas sobre a sua idade, sexo e se estava ou não familiarizado com as novas tecnologias. Nas subsecções seguintes mostraremos que resultados de satisfação foram obtidos consoante se tratava da utilização da aplicação, popularidade de utilizadores ou de grupos.

6.3.2 Tipos de utilizadores

Dos utilizadores que utilizaram a aplicação *GIA* e responderam ao questionário, 29% dos inquiridos são do sexo feminino enquanto que 71 % são do sexo masculino. Dos 48 inquiridos apenas 2 se apresentam como não conhecedores de novas tecnologias. No caso do uso de redes sociais, uma percentagem de 94% dos inquiridos, como podemos ver pela Figura 6.12, usa as redes sociais diariamente sendo que a mais utilizada pelos utilizadores é o *Facebook*. As faixas etárias dos utilizadores que responderam ao inquérito variam entre os 18 anos e os 51 anos.

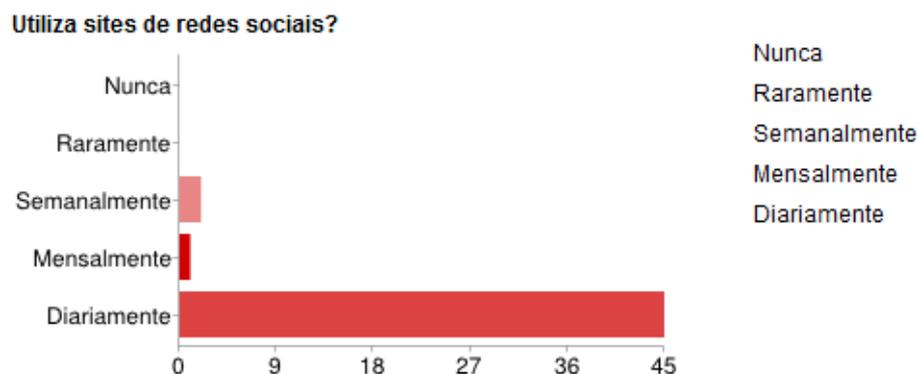


Figura 6.12: Utilização de redes sociais

6.3.3 Relevância de utilizadores

Sobre a popularidade de utilizadores para os grupos nas redes sociais, foram realizadas algumas perguntas tanto no âmbito da ideia em si, como também sobre a aplicação. Nestas perguntas poderiam classificar a resposta num intervalo de 1 a 5, em que 5 era o valor mais alto e o 1 o valor mais baixo, sendo o valor 3 considerado neutro.

No caso da questão de se um utilizador considera importante a ideia de classificarmos os utilizadores de um determinado grupo (pergunta nº3), a maioria dos utilizadores mostrou-se interessada com na ideia, com 60 % das respostas positivas e 25 % das respostas mais neutras, como podemos ver através da Figura 6.13. A pergunta 4 estava associada ao porquê de se gostar ou não da ideia de se calcular a popularidade dos utilizadores. Alguns utilizadores alegaram que não utilizavam os grupos nas redes sociais,

Considera importante o uso de uma ferramenta para indicar a relevância dos utilizadores para os grupos ?

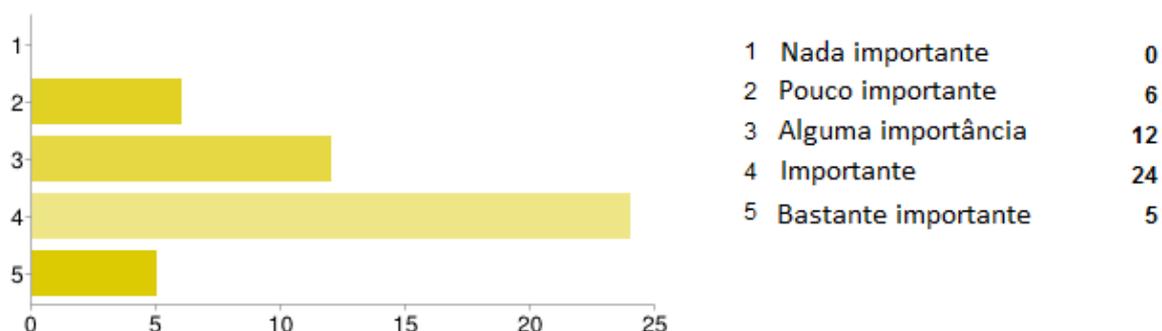


Figura 6.13: Interesse de calcular a popularidade de utilizadores

enquanto que outros acharam útil ter uma ferramenta que verifique que utilizadores é que participam de forma activa nos grupos.

Em relação aos dados mostrados pelo *GIA*, era questionado se os resultados da importância dos utilizadores para o grupo correspondiam ao que esperava. Os resultados obtidos foram bastante positivos, com 75% positivo, 19 % neutro e com apenas 6 % abaixo do esperado como se pode ver através da Figura 6.14 .

Em relação ao *GIA*, os resultados da importância dos utilizadores para o grupo correspondem ao que esperava?

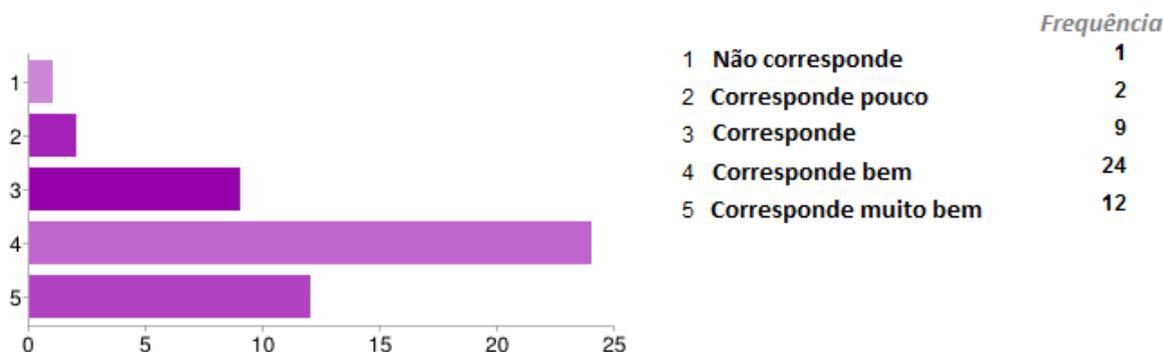


Figura 6.14: Interesse de uma ferramenta que verifique em que grupos um utilizador tem interesse

6.3.4 Relevância de grupos

No caso dos dados sobre a importância dos grupos para os utilizadores, a pergunta 5 questionava se daria importância a uma ferramenta que verificasse utilidade dos grupos nas redes sociais. Nesta pergunta 73% dos inquiridos demonstrou interesse positivo, 17 % neutro e 10 % não achou que existiria alguma utilidade nessa aplicação. A Figura 6.15 mostra os resultados obtidos nesta pergunta 5. Considera-se que é um bom indicador que as pessoas têm consciência que estão inscritas em demasiados grupos, em que muitas das

vezes não participam ou têm qualquer interesse.

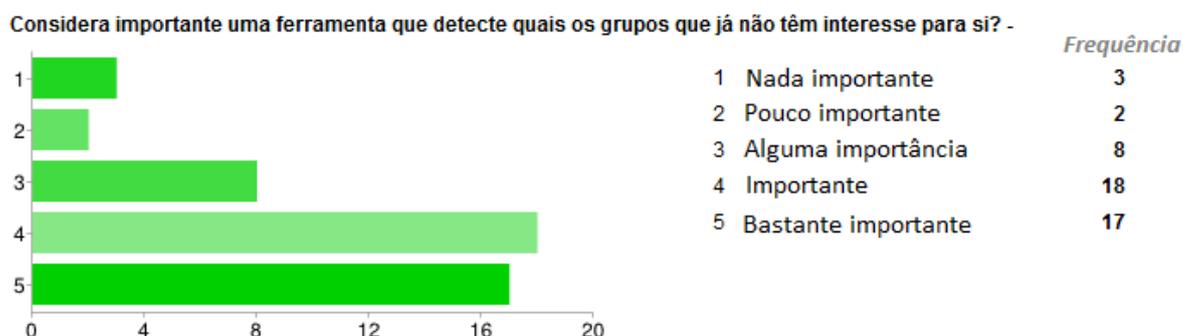


Figura 6.15: Interesse de uma ferramenta que verifique que grupos um utilizador tem interesse

Na pergunta 7, em que se perguntava a opinião sobre os valores obtidos pela aplicação *GIA* no âmbito da importância dos grupos para os utilizadores, 71% acham que os valores correspondem de forma positiva ao que se esperava, 23% de forma neutra e 6 % acham que os resultados estão abaixo do esperado. Este resultado é positivo pois foi conseguido que uma grande parte dos utilizadores ficasse satisfeito com estes valores e com a importância que os seus grupos pessoais têm para eles. Estes dados demonstram que as fórmulas propostas no modelo (Capítulo 3 e 4), se adaptam à maioria dos utilizadores. A resposta a esta pergunta por parte dos inquiridos, pode-se encontrar na Figura 6.16.

Em parte este bom resultado também se pode dever, ao comportamento competitivo dos utilizadores entre si, para verem quem é o utilizador com mais relevância para o grupo.

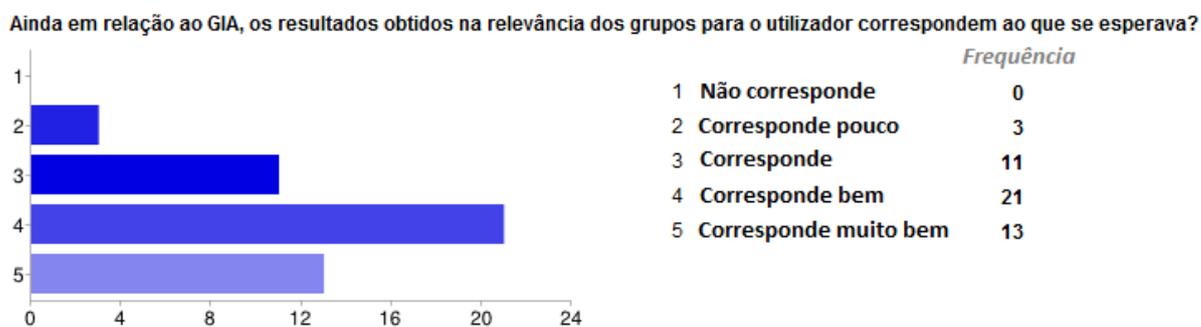


Figura 6.16: Popularidade dos grupos para o utilizador na aplicação *GIA*

6.3.5 Partilha de conteúdos para os grupos

No caso da partilha de conteúdos dos utilizadores para os grupos, os resultados foram um pouco diferentes. Na pergunta 8, questionava-se se ter uma funcionalidade que permita a partilha informação consoante os tópicos de cada grupo seria uma ferramenta interessante. Verificou-se que 63% dos inquiridos acham que era uma ferramenta interessante, 25 % acham que era indiferente e 12 % acham que não era interessante. Apesar dos números serem positivos, algumas das pessoas inquiridas não perceberam que vantagem é que essa funcionalidade poderia ter não só a nível de *marketing* mas também para partilhar informação apenas para os grupos cujo tema se adequava ao tema do *post*. A Figura 6.17 representa as respostas dadas ao questionário na pergunta 8.



Figura 6.17: Interesse numa funcionalidade de partilha de informação para os grupos consoante o tema do grupo

No caso da questão 9 onde se perguntava se a aplicação fazia bem a correspondência entre os *posts* e os grupos, foi onde se obtiveram resultados aparentemente piores. Na Figura 6.18 observa-se que 55% dos inquiridos respondeu positivamente, 33% neutralmente e 12% negativamente. Consideramos que estes resultados deveram-se ao facto de muitos grupos não terem descrição ou um nome suficientemente composto para se perceber qual o tema do grupo. Quando tal acontecia, não havia dados para fazer a correspondência entre os *posts* e os grupos. Não havendo dados, muitos dos inquiridos responderam de forma neutra, pois não tinham como avaliar se a ferramenta estava a fazer bem ou não as correspondências de forma correcta. Alguns utilizadores também não costumam partilhar conteúdos com frequência, o que também poderia não ajudar a sugerir *posts* para serem partilhados, tendo em conta que aceitar amigos ou jogar no *Facebook* não são considerados conteúdos para partilhar. Outro problema que surge em relação aos próprios *posts*, está relacionado com a dimensão dos *posts*. Se um *post* for demasiado pequeno, por exemplo menos 15 caracteres [Alc12c] e não tiver nenhum endereço associado, a extracção de tópicos a partir dele será difícil.

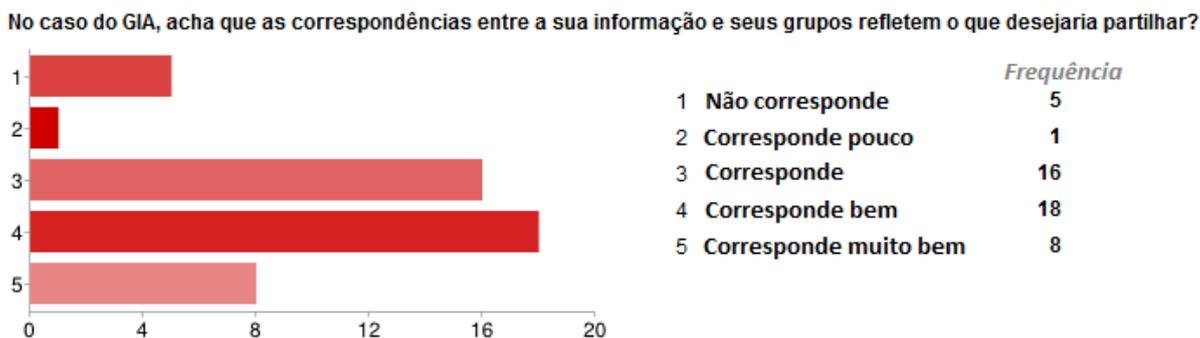


Figura 6.18: Avaliação da partilha de informação na aplicação GIA

6.3.6 Sugestões da parte dos utilizadores

No questionário realizado aos utilizadores, existiam algumas perguntas de resposta aberta, como as perguntas 12 e 13.

Na pergunta 12 sobre as novas funcionalidades sugeridas pelos utilizadores, surgiram várias respostas interessantes, tais como:

- *Seria interessante, mediante os posts que um utilizador possui no seu mural, sugerir uma lista de grupos onde o utilizador ainda não se encontra inscrito, de modo a este se poder inscrever em grupos que possam ter interesse para ele, mas que até então desconhecia;*
- *Reparei que utilizadores pertencentes ao grupo com um perfil privado e/ou wall bloqueada não são reconhecidos pela ferramenta.*
- *Análise de picos de actividade pelas alturas do dia/semana/mês;*
- *Criar opção que possibilite exportar os dados para excel;*
- *No separador "Share information" – possibilitar a edição do texto a publicar nos grupos;*
- *Criar uma opção para agendamento das publicações nos grupos (definir data e hora para ser feita a publicação automaticamente);*
- *Que sugerisse novos grupos de acordo com os grupos que o utilizador já aderiu.*

Na pergunta 13, em que se perguntava que outros dados seria interessante que a aplicação mostrasse, as respostas obtidas foram:

- *Talvez os "Gostos" de páginas do Facebook;*
- *Que fosse possível ver a minha relevância pessoal nos grupos de forma destacada.*
- *Número de visitas diárias do utilizador;*
- *Informação relativa ao tempo gasto na rede social ou numero de horas que o utilizador esteve on-line no chat de cada grupo;*

6.4 Resumo

Estes resultados são favoráveis à aprovação do modelo proposto, tendo ainda contribuído para identificar algumas limitações. Concluímos que os grupos mais pequenos, em geral, têm uma maior actividade, com o maior número de participantes que os médios ou grandes. Também verificamos, que apesar de existirem mais utilizadores a terem uma relevância nula em grupos pequenos, como existe mais partilhas seria de esperar que esse número aumentasse. O número de grupos com relevância nula é bastante elevado o que demonstra que muitos utilizadores estão inscritos em grupos, apesar da sua participação ser nula. Muitos grupos poderiam deixar de existir o que poderia permitir uma limpeza no número de grupos que o sistema dispõe. Esta informação pode sugerir que todos os sistemas como as redes sociais deviam ter mecanismos que detectassem a actividade dentro dos grupos e que ao fim de um determinado os eliminasse. Visto que o modelo proposto é bastante versátil, tal permitirá realizar vários melhoramentos futuros. Em relação ao inquéritos realizados aos utilizadores da aplicação *GIA*, algumas das sugestões podem influenciar o trabalho futuro, como por exemplo uma opção de agendamento de publicações para os grupos, pois iria facilitar utilizadores que quisessem fazer, por exemplo, publicidade para os grupos. A opção de se poder editar os *posts* do utilizadores antes de envia-los para os grupos é uma funcionalidade de bastante interesse pois poderia permitir uma personalização da informação para cada grupo, o que poderia melhorar a popularidade de cada conteúdo partilhado. No entanto existem algumas sugestões propõe funcionalidades que já existentes. Considera-se, que, este modelo se encontra no bom caminho, com os resultados obtidos em ambas as vertentes e aprovados por um conjunto de utilizadores inquiridos.



Conclusões e trabalho futuro

7.1 Conclusões

Neste trabalho propôs-se uma aplicação que permite que seja calculada a relevância da informação partilhada no contexto de grupos em redes sociais, avaliando também que utilizadores são mais importantes para cada grupo e quais os grupos que são efectivamente mais importantes para cada utilizador. Esta ferramenta também permite que um utilizador propague informação para os grupos, com base no tema do grupo e no tema associado ao conteúdo, de forma a propagar informação que realmente poderá interessar aos membros do grupo.

Antes do início do processo de concepção da aplicação, existiu uma primeira fase de estudo preliminar do tema, na qual se realizou um levantamento do estado da arte, após o que se procedeu a uma segunda fase, dedicada à elaboração da aplicação proposta.

Estes objectivos foram cumpridos, abrangendo múltiplas dimensões, desde a definição do conceito de relevância assente na quantificação de métricas de popularidade, considerado adequado aos contextos de utilização de grupos em redes sociais, passando depois pela concepção de um modelo e metodologia que permitissem capturar aquele conceito, seguido do desenho arquitectural e correspondente implementação num protótipo, com a devida aproximação ao caso real considerado, para a rede social *Facebook*.

Antes de se implementar uma ferramenta que tornasse possível uma propagação filtrada da informação, existiu a necessidade de se formular um modelo em que fosse possível calcular quantitativamente qual era o valor, na aceção de relevância considerada, de um determinado conteúdo num grupo.

Foi necessário também propor uma arquitectura e implementação que permitisse não

só a interacção com o *Facebook*, mas também com o *Alchemy*, extractor de conteúdos utilizado para analisar as informações dos grupos e também dos conteúdos partilhados pelos utilizadores.

Como resultado, a aplicação/ferramenta desenvolvia consegue identificar para os diferentes grupos, quais os utilizadores que partilham a informação que mais interessará aos membros do grupo. A aplicação desenvolvida permitiu concluir que o comportamento em grupos de pequenas dimensões difere, não só a nível da partilha, pois existem mais utilizadores distintos a partilhar nos grupos mais pequenos, mas também a nível de participação nos comentários e anotações aos conteúdos de grupos de média e grande dimensão.

Conclui-se também que a aplicação facilita a identificação dos vários grupos inactivos, aos quais os utilizadores poderiam deixar de pertencer por já não terem interesse neles.

A funcionalidade que suporta a propagação de informação foi implementada e concluída com sucesso. Consideramos, no entanto, que esta funcionalidade não foi ainda completamente explorada, uma vez que os perfis de grupo, no ambiente proposto, não tinham muita informação a partir da qual fosse possível concluir qual era o âmbito para o qual cada grupo havia sido criado.

Os resultados dos inquéritos aos utilizadores que participaram na validação desta aplicação foram bastante positivos, pois no geral os utilizadores evidenciaram um claro interesse pela aplicação e pelos seus objectivos, vendo nesta aplicação uma ferramenta com usabilidade. A dinâmica que a própria aplicação induziu no comportamento dos utilizadores, ao providenciar-lhes *feedback* sobre os seus níveis relativos de popularidade, é certamente um aspecto merecedor de atenção em desenvolvimentos futuros.

Ainda no âmbito da elaboração deste trabalho, foi submetido e aceite um artigo na conferencia internacional da especialidade, *The 2nd International Conference on Social Computing and its Applications (SCA 2012)* [LMC12].

Embora o núcleo do trabalho proposto possa ter terminado, existem várias modificações e melhoramentos que podem ser feitos, como se discute na Secção 7.2.

7.2 Trabalho futuro

Esta secção contempla algumas ideias sobre o trabalho futuro que poderá ser desenvolvido no âmbito do tema desta dissertação. Existem alguns ajustes que podem ser feitos, tanto a nível da implementação como a nível dos próprios conceitos conforme se descreve em seguida, segundo as diversas áreas em que se inserem. Na área dos conceitos e na forma como as relevâncias são calculadas, um melhoramento seria contemplar utilizadores que são mais precisos/eficientes na sua partilha de informação. Por exemplo, um utilizador que partilhe 3 *posts* e se existirem 3 utilizadores que gostam dessas 3 publicações no total terá 9 gostos o que é diferente de um utilizador que apenas numa publicação consegue logo que 9 utilizadores gostem do que ele partilhou. Nesse caso, o utilizador

mais preciso/eficiente deveria ter mais relevância do que o menos preciso/eficiente, apesar de o número de gostos de ambos os utilizadores serem os mesmos.

Usando o conceito das relações entre os utilizadores que partilham a informação e aqueles que comentam ou gostam da informação, pode-se verificar se as interações que acontecem no grupo são devido ao grau de afinidade (força das ligações) entre utilizadores ou porque a informação é realmente importante. No caso de ser apenas por afinidade, essa relevância da informação partilhada deve descer de forma significativa para não ser confundida com relevância da informação com afinidade entre utilizadores. Sobre a forma como esse grau de afinidade é calculado, já existem vários trabalhos nessa área, como verificar para quem é que um utilizador envia mensagens privadas ou públicas e com que frequência. Poder-se-ia, desta forma completar o trabalho já feito com esta nova preocupação e assim refinar os conceitos encontrados nas fórmulas propostas. Poder-se-ia, ainda perceber qual o grau de afinidade entre os conjuntos de utilizadores no caso em que apenas 3 ou 4 utilizadores participam num determinado grupo e sugerir a criação de um grupo mais pequeno, com esses utilizadores como membros.

Na área da relevância da informação e analisando a nível geral dos utilizadores, poderia ser criada uma classificação de notícias mais populares, partilhadas através da aplicação e uma lista de grupos com os grupos com maior relevância encontrados. Este tipo de informação poderia provocar uma certa competição entre grupos e dessa forma estimular a partilha de informação. Esta análise também poderia verificar que tipo de notícias tinham mais relevância e permitir assim que estes tipos de conceitos pudessem ajudar a fazer estudos de mercado que poderiam ajudar em áreas como o *marketing*, publicidade e relações públicas. Ainda nesta área poderia-se utilizar as duas partes distintas da aplicação e avaliar os *posts* que eram partilhados através da aplicação. Esses *posts* partilhados pelo *GIA* seriam analisados e a sua relevância calculada e assim verificando se as sugestões de partilha da informação estavam correctas ou se havia ajustes a serem feitos.

Para finalizar os novos melhoramentos para trabalho futuro, surge a área da análise de tópicos encontrados nos *posts* dos utilizadores por forma a sugerir dentro das temáticas, vários grupos que poderiam interessar ao utilizador. Apesar de existir uma ferramenta no *Facebook* que sugere grupos, ela baseia-se no números de amigos que há em comum naquele grupo com o utilizador e não nos temas que esse utilizador agora tem partilhado. O facto desta ferramenta se adequar ao tema em que agora o utilizador está interessado, demonstra a maior versatilidade que esta ferramenta pode vir a ter, pois pode-se adaptar aos vários temas de interesse, a curto ou a longo prazo.

Bibliografia

- [AA04] R. Almeida e V. Almeida. A community-aware search engine. In *Proceeding of the 2004 World Wide Web Conference, 2004*.
- [ABC⁺12] Thies C. Arntzen, Stig Bakken, Shane Caraveo, Andi Gutmans, Rasmus Lerdorf, Sam Ruby, Sascha Schumann, Zeev Suraski, Jim Winstead, e Andrei Zmievski. PHP. Internet <http://www.php.net/>, Setembro 2012. Consultado a 10/09/2012.
- [Act12] Activision. Call of Duty. Internet <http://www.callofduty.com/>, Setembro 2012. Consultado em 15/09/2012.
- [Aja12] Ajax. Internet <http://api.jquery.com/jquery.ajax/>, Setembro 2012. Consultado a 10/09/2012.
- [Alc12a] AlchemyAPI - Linguas detectaveis. Internet <http://www.alchemyapi.com/api/lang/>, Setembro 2012. Consultado em 08/09/2012.
- [Alc12b] Alchemy. AlchemyAPI. Internet <http://www.alchemyapi.com/>, Fevereiro 2012. Consultado em 03/02/012.
- [Alc12c] Alchemy. AlchemyAPI - Get Text language. Internet <http://www.alchemyapi.com/api/lang/textc.html>, 05 2012. Consultado em 08/05/2012.
- [Alt12] Altimeter Group. Internet <http://www.altimetergroup.com/>, Setembro 2012. Consultado em 15/09/2012.
- [Ama12] Amazon. Amazon. Internet <http://www.amazon.com/>, Fevereiro 2012. Consultado em 06/01/2012.
- [Ato12] Atom- Definition. Internet <http://www.atomenabled.org/>, Setembro 2012. Consultado em 08/09/2012.

- [Aut12a] Auth 2.0 - oauth 2.0 is the next evolution of the oauth protocol which was originally created in late 2006. Internet <http://oauth.net/2/>, Setembro 2012. Consultado em 16/09/2012.
- [Aut12b] Auth 2.0 facebook. Internet <https://developers.facebook.com/docs/authentication/>, Setembro 2012. Consultado em 16/09/2012.
- [AYFM99] M. N. Ahmed, S. M. Yamany, A. A. Farag, e T. Moriarty. Bias field estimation and adaptive segmentation of mri data using a modified fuzzy c-means algorithm. In *In Proc. IEEE Int. Conf. Computer Vision and Pattern Recogn*, pág. 250–255, 1999.
- [BE07] Danah. Boyd e Nicole B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.
- [Bli12] Blizzard. World of Warcraft. Internet us.battle.net/wow/pt/, Setembro 2012. Consultado em 15/09/2012.
- [Boy08] Danah Boyd. *Taken Out of Context: American Teen Sociality in Networked Publics*. Tese de Doutorado, University of California-Berkeley, School of Information., 2008.
- [BRAYC⁺10] Senjuti Basu Roy, Sihem Amer-Yahia, Ashish Chawla, Gautam Das, e Cong Yu. Space efficiency in group recommendation. *The VLDB Journal*, 19(6):877–900, Dezembro 2010.
- [BSS⁺08] Shumeet Baluja, Rohan Seth, D. Sivakumar, Yushi Jing, Jay Yagnik, Shankar Kumar, Deepak Ravichandran, e Mohamed Aly. Video suggestion and discovery for youtube: taking random walks through the view graph. In *Proceedings of the 17th international conference on World Wide Web, WWW '08*, pág. 895–904, New York, NY, USA, 2008. ACM.
- [Bur07] Robin D. Burke. Hybrid web recommender systems. In Peter Brusilovsky, Alfred Kobsa, e Wolfgang Nejdl, editores, *The Adaptive Web, Methods and Strategies of Web Personalization*, volume 4321 of *Lecture Notes in Computer Science*, pág. 377–408. Springer, 2007.
- [Cas00] M. Castells. *The Rise of The Network Society: The Information Age: Economy, Society and Culture*. Information Age. John Wiley & Sons, 2000.
- [CBH02] A. Crossen, J. Budzik, e K. Hammond. Flytrap: Intelligent group music recommendation. In *Gil, Y., Leake, D., eds.: IUI 2002: International Conference on Intelligent User Interfaces.*, ACM, New York (2002) 184-185:2, 2002.
- [CG08] Edward Curry e Paul Grace. Flexible self-management using the model-view-controller pattern. *IEEE Software*, 25(3):84–90, May 2008.

- [Cit12] CiteULike. Citeulike. Internet <http://en.wikipedia.org/wiki/CiteULike>, Setembro 2012. Consultado a 13/09/2012.
- [CLT11] Juan J. Cameron, Carson Kai-Sang Leung, e Syed K. Tanbeer. Finding strong groups of friends among friends in social networks. *Dependable, Autonomous and Secure Computing, IEEE International Symposium on*, 0:824–831, 2011.
- [com12] Facebook company. Graph api facebook. Internet <https://developers.facebook.com/docs/reference/api/>, Setembro 2012. Consultado em 16/09/2012.
- [Con12a] Connontea - organizar, partilhar e descobrir. Internet <http://www.connotea.org/>, Setembro 2012. Consultado a 14/09/2012.
- [Con12b] Content analysis - concepts. Internet http://en.wikipedia.org/wiki/Content_analysis, Setembro 2012. Consultado em 15/09/2012.
- [CQL08] Daniel Cutting, Aaron Quigley, e Bjarn Landfeldt. Spice: Scalable p2p implicit group messaging. *Computer Communications*, 31(3):437 – 451, 2008.
- [CSP11] Kevin R. Canini, Bongwon Suh, e Peter L. Piroli. Finding credible information sources in social networks based on content and social structure. In *Proceedings of the third IEEE International Conference on Social Computing (SocialCom)*, 2011.
- [cUR12] cURL. cURL- Client for URLs. Internet <http://curl.haxx.se/>, Maio 2012. Consultado em 07/05/2012.
- [Emp12] Empresa amCharts . amCharts. Internet <http://www.amcharts.com/>, Setembro 2012. Consultado em 10/09/2012.
- [Enz01] Matisse Enzer. Why the Internet Will Never Die. <http://www.matisse.net/opinion/internet-alive.html>, Dezembro 2001. Consultado em 30/01/2012.
- [Est96] Martin Ester. A density-based algorithm for discovering clusters in large spatial databases with noise. pág. 226–231. AAAI Press, 1996.
- [Fac11a] Facebook. About Facebook Plataform. Internet <http://www.facebook.com/facebook?sk=info>, Outubro 2011. Consultado em 02/10/2011.
- [Fac11b] Facebook. Facebook Plataform Policies. <http://developers.facebook.com/policy/>, Outubro 2011. Consultado em 20/10/2011.
- [Fac11c] Facebook. SDKs- Facebook. <http://developers.facebook.com/docs/sdks>, Novembro 2011. Consultado em 16/11/2012.

- [Fac12a] Facebook. Facebook - deprecated share button. Internet https://www.facebook.com/note.php?note_id=10150114712634716, Maio 2012. Consultado em 07/05/12.
- [Fac12b] Facebook. Facebook - HTTPS -SSL. Internet <https://developers.facebook.com/blog/post/497/>, Maio 2012. Consultado em 07/05/2012.
- [Fac12c] Facebook. Facebook API - User. Internet <https://developers.facebook.com/docs/reference/api/user/>, Setembro 2012. Consultado em 09/09/2012.
- [Fac12d] Facebook. Facebook Applications. Internet - <https://developers.facebook.com/docs/guides/canvas/>, Julho 2012. Consultado em 15/07/2012.
- [Fac12e] Facebook. Facebook Insights. Internet <https://developers.facebook.com/docs/guides/canvas/insights>, July 2012. Consultado em 15/07/2012.
- [Fac12f] Facebook. Facebook Recommendation Plugin. <http://developers.facebook.com/docs/reference/plugins/recommendations2/>, Janeiro 2012. Consultado em 31/01/2012.
- [Fac12g] Facebook. Facebook Suggesting Friends. http://www.facebook.com/help/friends/people_you_may_know, Janeiro 2012. Consultado em 31/01/2012.
- [Fac12h] Facebook. Funcionalidade Seen by - Facebook. Internet <http://newsroom.fb.com/News/Update-to-Facebook-Groups-186.aspx>, Julho 2012. Consultado em 01/09/2012.
- [Fac12i] Facebook Company. Facebook PHP SDK. Internet <https://developers.facebook.com/docs/reference/php/>, Setembro 2012. Consultado em 16/09/2012.
- [Fan12] Fancy Box. Internet <http://fancybox.net/>, Setembro 2012. Consultado a 10/09/2012.
- [FGM⁺99] Roy T. Fielding, James Gettys, Jeffrey C. Mogul, Henrik Frystyk Nielsen, Larry Masinter, Paul J. Leach, e Berners-Lee. Hypertext transfer protocol – http/1.1. Internet <http://tools.ietf.org/html/rfc2616>, Julho 1999. Consultado em 08/09/2012.
- [Fit99] Brad Fitzpatrick. Live Journal, Abril 1999. Consultado em <http://www.livejournalinc.com/aboutus.php>.

- [Fli12a] Flickr. Flickr. Internet url<http://www.flickr.com/>, May 2012. Consultado a 15-05-2012.
- [Fli12b] Flickr. Flickr apis. Internet <http://www.flickr.com/services/api/>, Setembro 2012. Consultado em 19/09/2012.
- [Fot12] Fotolog. Fotolog. Internet <http://www.fotolog.com/a/info/aboutus>, Fevereiro 2012. Consultado em 03/02/2012.
- [Fow03] Martin Fowler. *Patterns of Enterprise Application Architecture*. Pearson Education, Inc, 2003.
- [Fri12] Friendster. Friendster. Internet <http://www.molglobal.net/>, Fevereiro 2012. Consultado em 03/02/2012.
- [GK09] Eric Gilbert e Karrie Karahalios. Predicting tie strength with social media. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pág. 211–220, New York, NY, USA, 2009. ACM.
- [Gof59] Erving Goffman. *Presentation of self in everyday life*. Doubleday Anchor Books, 1959.
- [Goo] Google. Historia Google+. <http://www.google.com/intl/en/about/corporate/company/history.html>. Consultado em 2011.11.14.
- [Goo11a] Google. Funcionalidades Google+. <http://www.google.com/intl/pt-PT/+learnmore/>, Novembro 2011. Consultado em 14/11/2011.
- [Goo11b] Google. Google Mail. <https://mail.google.com/mail>, Novembro 2011.
- [Goo12a] Google. Google Survey. Internet <http://www.google.com/insights/consumersurveys/home>, Agosto 2012. Consultado em 08/08/2012.
- [Goo12b] Google. Sobre YouTube. Internet http://www.youtube.com/t/about_youtube, Setembro 2012. Consultado em 08/09/2012.
- [Gra73] Mark Granovetter. The impact of social structure on economic outcomes. *American Journal of Sociology*, Vol.78:pp. 1360–1380, 1973.
- [Htt12] Https. Https - Definicao e o seu uso. Internet <https://www.eff.org/https-everywhere/faq>, July 2012. Consultado em 17/07/2012.
- [IMD12] IMDB. Imdb. Internet <http://www.imdb.com>, Setembro 2012. Consultado a 13/09/2012.

- [Inn12] Inno DB - Definition. Internet <http://www.innodb.com/>, July 2012. Consultado em 17/07/2012.
- [JFv12] Kien La Jacques Fuentes e Clay vanSchalkwijk. Active Record PHP. <http://www.phpactiverecord.org/>, Maio 2012. Consultado em 28-05-2012.
- [Jqu12] JQuery -Library. Internet <http://jquery.com/>, Setembro 2012. Consultado a 10/09/2012.
- [JS07] Anthony Jameson e Barry Smyth. Recommendation to groups. In Peter Brusilovsky, Alfred Kobsa, e Wolfgang Nejdl, editores, *The Adaptive Web: Methods and Strategies of Web Personalization*, pág. 596–627. Springer, Berlin, 2007.
- [JSOto] Json - Definition. Internet <http://www.json.org/>, 2012 Agosto. Consultado em 30/08/2012.
- [KMN⁺02] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, Angela Y. Wu, Senior Member, e Senior Member. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:881–892, 2002.
- [KP88] Glenn E. Krasner e Stephen T. Pope. A cookbook for using the model-view controller user interface paradigm in smalltalk-80. *J. Object Oriented Program.*, 1(3):26–49, Agosto 1988.
- [Kru97] Bruce Krulwich. Lifestyle finder: Intelligent user profiling using large-scale demographic data. *AI Magazine*, 18(2):37–45, 1997.
- [Lim12] Facebook Limit. Internet <https://developers.facebook.com/blog/post/478/>, Setembro 2012. Consultado em 16/09/2012.
- [Lin11] LinkedIn. LinkedIn APIs. <https://developer.linkedin.com>, Novembro 2011. Consultado em 16/11/2011.
- [Lin12a] LinkedIn. About LinkedIn. <http://br.press.linkedin.com/about>, Fevereiro 2012. Consultado em 03/02/2012.
- [Lin12b] LinkedIn. LinkedIn Groups. http://www.linkedin.com/createGroup?displayCreate=&trk=hb_side_crgrp, Janeiro 2012. Consultado em 31/01/2012.
- [Lin12c] LinkedIn. Suggesting People LinkedIn. <https://developer.linkedin.com/documents/people-search-api>, Janeiro 2012. Consultado em 31/01/2012.

- [LMC12] Tânia Leitão, Carmen Morgado, e José Cardoso Cunha. Measuring Popularity In Social Network Groups. *SCA 2012: International Conference on Social Computing and its Applications*, 2012.
- [LVDV99] Henry Lieberman, Neil W. Van Dyke, e Adrian S. Vivacqua. Let's browse: a collaborative web browsing agent. In *Proceedings of the 4th international conference on Intelligent user interfaces, IUI '99*, pág. 65–68, New York, NY, USA, 1999. ACM.
- [McC02] Joseph F. McCarthy. Pocket restaurant finder: A situated recommender systems for groups. In *Proceeding of Workshop on Mobile Ad-Hoc Communication at the 2002 ACM Conference on Human Factors in Computer Systems*, 2002.
- [MCCC09] Carmen Morgado, José C. Cunha, Jorge F. Custódio, e Nuno Correia. A group-based model for dynamic communities. In *The International Conference on Computational Aspects of Social Networks (CASoN 2009), Fontainebleau, France, June 24-27, IEEE Computer Society 2009*, pág. 161–164, 2009.
- [Men12] Leonardo Mendes. AlojamentoVivo - Facebook. Internet <https://www.facebook.com/Alojamentovivo>, Setembro 2012. Consultado em 17/09/2012.
- [Min10] Miniwatts. Internet World Stats. <http://internetworldstats.com/stats.htm>, Janeiro 2010. Consultado em 28/01/2010.
- [MKR04] Bradley N. Miller, Joseph A. Konstan, e John Riedl. Pocketlens: Toward a personal recommender system. *ACM Trans. Inf. Syst.*, 22:437–476, July 2004.
- [MMG⁺07] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, e Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, IMC '07*, pág. 29–42, New York, NY, USA, 2007. ACM.
- [Moo12] Moodle. Moodle - conceito. Internet <http://moodle.org/>, Setembro 2012. Consultado em 10/09/2012.
- [MyI12] Myisam- definicao. Internet <http://dev.mysql.com/doc/refman/5.0/en/myisam-storage-engine.html>, July 2012. Consultado 17/07/2012.
- [MyS12a] MySpace. Internet <http://www.myspace.com/>, Fevereiro 2012. Consultado a 03/02/2012.
- [MyS12b] MySql - Definition. Internet <http://www.mysql.com/>, July 2012. Consultado em 17/07/2012.

- [NRW12] Professor of History Norman R. West. Human groups in history. Internet <http://www2.sunysuffolk.edu/westn/humangroups.html>, Janeiro 2012. Consultado em 23/01/2012.
- [OAu12] OAuth- An open protocol to allow secure authorization in a simple and standard method from web, mobile and desktop applications. Internet <http://oauth.net/>, Setembro 2012. Consultado em 08/09/2012.
- [Ope12a] OpenCalais. OpenCalais. <http://www.opencalais.com/about>, Fevereiro 2012. Consultado em 03/02/2012.
- [Ope12b] OpenCalais. OpenCalais Web Service. Internet <http://www.opencalais.com/>, 2012. Consultado em 04/02/2012.
- [Pai11] Miguel Rangel Pais. Grupos implícitos em aplicações interactivas. Tese de Mestrado, Engenharia Informática - Faculdade de Ciência e Tecnologia - Universidade Nova de Lisboa, Setembro 2011.
- [Pal10] T. Palsma. Discovering groups using short messages from social network profiles. Tese de Mestrado, University of Twente, July 2010.
- [Pet12] PeterVanDijck. Xfml. Internet <http://www.cmswiki.com/tiki-pagehistory.php?page=XFML&diff=2>, Setembro 2012. Consultado em 16/09/2012.
- [Pik12] Rob 'Commander' Pike. Utf8. <http://www.cl.cam.ac.uk/~mgk25/ucs/utf-8-history.txt>, July 2012. Consultado em 17/07/2012.
- [PM10] Ralph Meijer Peter Millard, Peter Saint-Andre. Xep-0060: Publish-subscribe. *The 8th Conference of Open Innovation Framework Program FRUCT*, 2010.
- [PMC11] Miguel Rangel Pais, Carmen Morgado, e José C. Cunha. Implicit groups in web-based interactive applications. In *CASoN -International Conference on Computational Aspects of Social Networks, CASoN 2011, Salamanca, Spain, October 19-21, 2011*, pág. 175–180, 2011.
- [Pre12] Oxford University Press. Oxford Dictionaries - The world most trusted dictionaries. <http://oxforddictionaries.com/definition/group?q=group>, Janeiro 2012. Consultado em 20/01/2012.
- [Pro09] Pew Research Centeris Internet American Life Project. Adults on social network sites. <http://www.pewinternet.org/Infographics/Growth-in-Adult-SNS-Use-20052009.aspx> (2009), 1:1, 2009.
- [Pyt12] Python. Internet <http://www.python.org/>, Setembro 2012. Consultado em 15/09/2012.

- [Que10] Paulo Querido. Estar em rede não é um dever nem uma obrigação, mas confere uma vantagem . <http://pauloquerido.pt/media/>, Junho 2010. Consultado em 20/01/2012.
- [Rai12] Wiki raider. Internet <http://www.wikiraider.com/index.php/>, Setembro 2012. Consultado a 15/09/2012.
- [RBDD⁺10] Maayan Roth, Assaf Ben-David, David Deutscher, Guy Flysher, Ilan Horn, Ari Leichtberg, Naty Leiser, Yossi Matias, e Ron Merom. Suggesting friends using the implicit social graph. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '10*, pág. 233–242, New York, NY, USA, 2010. ACM.
- [RDF12] Rdf - conceito. Internet <http://www.w3.org/RDF/>, Setembro 2012. Consultado em 15/09/2012.
- [REL12] Rel-tag. Internet <http://microformats.org/wiki/rel-tag>, Setembro 2012. Consultado em 15/09/2012.
- [Ric06] R. Richards. *Representational State Transfer (REST)* in *Pro PHP XML and Web Services*, chapter 17, pág. 633–672. Springer Publishing, 2006.
- [RSS12] RSS. Internet <http://www.rssboard.org/rss-0-9-0>, Setembro 2012. Consultado em 08/09/2012.
- [SA06] James Salter e Nick Antonopoulos. Cinemascreen recommender agent: Combining collaborative and content-based filtering. *IEEE Intelligent Systems*, 21(1):35–41, 2006.
- [SA09] Silvia Schiaffino e Analía Amandi. Artificial intelligence. chapter Intelligent user profiling, pág. 193–216. Springer-Verlag, Berlin, Heidelberg, 2009.
- [Sch09] Dan Schawbel. Build a Marketing Platform Like a Celebrity. On Business week, Agosto 2009. Consultado em 20/01/2012.
- [Sec12] Second Life. Internet <http://secondlife.com/>, Setembro 2012. Consultado em 15/09/2012.
- [Sil11] Anibal Cavaco Silva. Perfil Facebok Anibal Cavaco Silva. Internet <http://www.facebook.com/CavacoSilva>, Outubro 2011. Consultado em 26/10/2011.
- [Sim12] Simcity social Facebook. Internet <https://www.facebook.com/simcitysocial>, Setembro 2012. Consultado em 17/09/2012.
- [Six12] Sixdegrees. Internet <http://www.sixdegrees.org/about>, Setembro 2012. Consultado em 03/02/2012.

- [SNCA⁺09] Elizeu Santos-Neto, David Condon, Nazareno Andrade, Adriana Iamnit-chi, e Matei Ripeanu. Individual and social behavior in tagging systems. In *Proceedings of the 20th ACM conference on Hypertext and hypermedia, HT '09*, pág. 183–192, New York, NY, USA, 2009. ACM.
- [Sne06] James Snell. Getting to know the atom publishing protocol, part 1: Create and edit web resources with the atom publishing protocol. Internet <http://www.ibm.com/developerworks/library/x-atomppl/>, 2006. Consultado em 10/09/2012.
- [SOA12] SOAP. Soap - soap is a simple xml-based protocol to let applications exchange information over http. Internet <http://www.w3schools.com/soap/default.asp>, Setembro 2012. Consultado em 16/09/2012.
- [SPUP02] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, e David M. Pen-nock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and develop-ment in information retrieval, SIGIR '02*, pág. 253–260, New York, NY, USA, 2002. ACM.
- [SRGDA11] Lara Quijano Sanchez, Juan A. Recio-García, e Belén Díaz-Agudo. Happy-Movie: A Facebook Application for Recommending Movies to Groups. In *ICTAI - International Conference on Tools with Artificial Intelligence*, pág. 239–244, 2011.
- [Ste12] Steam. Steam. <http://store.steampowered.com/recommended/>, Janeiro 2012. Consultado em 24/01/2012.
- [TD06] Aravind Srinivasan Thomas DuBois, Jennifer Golbeck. Rigorous probabi-listic trust-inference with applications to clustering. *University of Maryland*, pág. 4, 2006.
- [TMB09] Jaime Teevan, Meredith Ringel Morris, e Steve Bush. Discovering and using groups to improve personalized search. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining, WSDM '09*, pág. 15–24, New York, NY, USA, 2009. ACM.
- [TR10] Dario Taraborelli e Camille Roth. *Viable Web communities: Two case studies*. Kluwer, 2010.
- [TSK05] Pang-Ning Tan, Michael Steinbach, e Vipin Kumar. *Introduction to Data Mi-ning, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Bos-ton, MA, USA, 2005.
- [Twi11a] Twitter. Twitter - Definition. <http://twitter.com/about>, Novembro 2011. Consultado em 14/11/2011.

BIBLIOGRAFIA

- [Twi11b] Twitter. Twitter API FAQ. <https://dev.twitter.com/docs/api-faq>, Outubro 2011. Consultado em 10/10/2011.
- [UK12] IGN Entertainment UK. Ign. Internet <http://uk.corp.ign.com/>, Setembro 2012. Consultado a 13/09/2012.
- [W3C12a] W3C. Semantic Web - Definition. Internet <http://www.w3.org/2001/sw/>, Setembro 2012. Consultado em 22/09/2012.
- [W3C12b] W3C. What is CSS? Internet <http://www.w3.org/standards/webdesign/htmlcss>, Setembro 2012. Consultado a 10/09/2012.
- [Wal12] Wall Facebook. Internet http://www.webopedia.com/TERM/F/Facebook_wall.html, Setembro 2012. Consultado em 08/09/2012.
- [XLC12] Haoran Xie, Qing Li, e Yi Cai. Community-aware resource profiling for personalized search in folksonomy. *J. Comput. Sci. Technol.*, 27(3):599–610, 2012.
- [XML12] XML- Definicao. Internet <http://www.w3.org/TR/REC-xml/>, Setembro 2012. Consultado em 08/09/2012.
- [You11] YouTube. Data APIs- YouTube. Internet <http://code.google.com/intl/pt-PT/apis/youtube/>, Novembro 2011. Consultado em 10/11/2011.
- [Zin12] Zinga. FarmVille Facebook. Internet <https://www.facebook.com/FarmVille>, Setembro 2012. Consultado em 17/09/2012.