

Universidade Federal do Estado do Rio de Janeiro
Departamento de Informática Aplicada
Programa de Pós-Graduação em Informática

UTILIZANDO CONTEÚDO SOCIAL PARA O ENRIQUECIMENTO SEMÂNTICO
DE DIRETÓRIOS DE DESCOBERTA E ESCOLHA DE PLATAFORMAS NA
ECONOMIA COMPARTILHADA

Alexandre de Oliveira Paixão

Orientador: Astério Kiyoshi Tanaka
Co-orientador: Kate Cerqueira Revoredo

Rio de Janeiro
Julho de 2017

UTILIZANDO CONTEÚDO SOCIAL PARA O ENRIQUECIMENTO SEMÂNTICO
DE DIRETÓRIOS DE DESCOBERTA E ESCOLHA DE PLATAFORMAS NA
ECONOMIA COMPARTILHADA

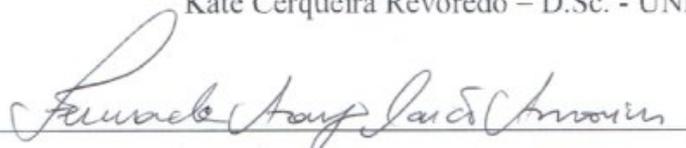
Alexandre de Oliveira Paixão

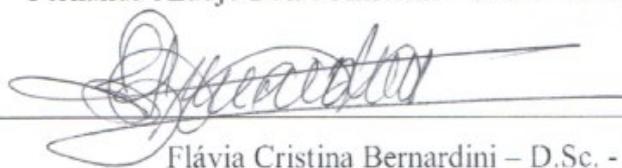
DISSERTAÇÃO APRESENTADA COMO REQUISITO PARCIAL PARA
OBTENÇÃO DO TÍTULO DE MESTRE PELO PROGRAMA DE PÓS-
GRADUAÇÃO EM INFORMÁTICA DA UNIVERSIDADE FEDERAL DO ESTADO
DO RIO DE JANEIRO (UNIRIO). APROVADA PELA COMISSÃO
EXAMINADORA ABAIXO ASSINADA.

Aprovada por:


Asterio Kiyoshi Tanaka – D.Sc. – UNIRIO


Kate Cerqueira Revoredo – D.Sc. - UNIRIO


Fernanda Araújo Baião Amorim – D.Sc. – UNIRIO


Flávia Cristina Bernardini – D.Sc. - UFF

RIO DE JANEIRO, RJ - BRASIL
Julho de 2017

Catálogo informatizado pelo(a) autor(a)

- P149 Paixão, Alexandre de Oliveira
Utilizando conteúdo social para o enriquecimento semântico de diretórios de descoberta e escolha de plataformas na economia compartilhada / Alexandre de Oliveira Paixão. -- Rio de Janeiro, 2017.
101 f.
- Orientador: Astério Kiyoshi Tanaka.
Coorientadora: Kate Cerqueira Revoredo.
Dissertação (Mestrado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Informática, 2017.
1. Análise de Sentimentos. 2. Mineração de Opiniões. 3. Economia Compartilhada. 4. Plataformas ponto-a-ponto de consumo e compartilhamento colaborativo. I. Tanaka, Astério Kiyoshi, orient. II. Revoredo, Kate Cerqueira, coorient. III. Título.

DEDICATÓRIA

“... se chamavam sonhos e sonhos não envelhecem...”

AGRADECIMENTOS

Agradeço, primeiramente, a Deus. Porque “Dele, por Ele e para Ele são todas as coisas”.

Agradeço à minha família: esposa, filhos e meus pais, que de forma tão amorosa e altruísta soube entender os meus longos finais de semana sozinho, voltado inteiramente aos estudos, sempre me apoiando e me incentivando a seguir adiante. Agradeço pelo chorar e sorrir juntos. Essa conquista é para vocês!

Agradeço aos meus amigos, aqueles que me incentivaram e animaram, e das mais diversas formas se fizeram presentes.

Agradeço aos professores da UNIRIO com os quais tive o privilégio de conviver e tanto aprender ao longo do Mestrado. Agradeço de forma especial à professora Kate Revoredo, que, com muita disponibilidade e zelo se dispôs a co-orientar esse trabalho, dando ao mesmo a direção que tanto necessitava.

Agradeço, por fim, aos meus colegas de mestrado, pelo convívio sempre respeitoso e inspirador.

Obrigado!

RESUMO

O crescimento constante da Economia Compartilhada (*Sharing Economy*) como fenômeno social e cultural tem causado mudanças significativas na sociedade atual e, sobretudo, nas relações de consumo, gerando um novo modelo de negócios onde se destaca o protagonismo dos indivíduos. Vemos ainda, acompanhando esse crescimento, as plataformas ponto-a-ponto (P2P) de compartilhamento e consumo colaborativo (SCC) se solidificarem como o meio tecnológico que viabiliza a partilha de recursos diretamente entre os indivíduos. Entretanto, o crescimento e também a diversidade das plataformas existentes levantou a questão de como se escolher a melhor entre as diferentes opções. Vindo ao encontro dessa questão foram propostas várias iniciativas, com destaque para os diretórios que permitem a pesquisa e a recuperação de plataformas. O diretório web Discover2Share é um exemplo dessas propostas, cuja particularidade é fazer uso de uma ontologia para descrever e recuperar as plataformas disponíveis com mais precisão, de acordo com as demandas dos usuários. No entanto, em muitas situações, o Discover2Share apresenta muitas opções, tornando difícil para o usuário escolher a melhor plataforma. Nessa dissertação é proposta uma abordagem que enriquece com opiniões, provenientes de conteúdo social, os dados existentes hoje no Discover2Share, tendo em vista ainda melhorar o processo de identificação da melhor plataforma de consumo e compartilhamento colaborativo. Como forma de avaliação foram realizados dois estudos de caso: um experimento considerando três plataformas (9Flats, Airbnb e HomeAway) que fornecem recursos do tipo ‘acomodações’ e com conteúdo social oriundo de três fontes (Twitter, Facebook e as próprias plataformas); e um questionário online para validar tal abordagem junto aos usuários do Discover2Share.

Palavras-chave: Análise de Sentimentos; Mineração de Opiniões; Economia Compartilhada; Plataformas ponto-a-ponto de Consumo e Compartilhamento Colaborativo, Processo de Tomada de Decisão.

ABSTRACT

The constant growth of the Sharing Economy as a social and cultural phenomenon has caused significant changes in the current society and, above all, in consumer relations, generating a new business model in which the protagonism of individuals is highlighted. Along with this growth, peer-to-peer (P2P) platforms for sharing and collaborative consumption (SCC) solidify as the technological environment that enables the sharing of resources directly between individuals. However, the growth and diversity of existing platforms has raised the question of how to choose the best among the different options. Coming to this issue, several initiatives were proposed, especially the directories that allow research and recovery of platforms. The Discover2Share web directory is an example of these proposals, whose particularity is to make use of an ontology to describe and retrieve the available platforms with more precision, according to the demands of the users. However, in many situations, Discover2Share has many options, making it difficult for the user to choose the best platform. In this dissertation it is proposed an approach that enriches with opinions, from social content, the available data currently in Discover2Share, in order to still improve the process of choosing the best platform of consumption and collaborative sharing. As an evaluation two case studies were carried: an experiment considering three platforms (9Flats, Airbnb and HomeAway) that provide 'accommodation' and social content resources from three sources (Twitter, Facebook and the platforms themselves); and an online survey to validate such an approach with Discover2Share users.

Keywords: Sentiment Analysis; Opinion Mining; Sharing Economy; Peer-to-Peer Sharing and Collaborative Consumption platforms; Decision-making process.

SUMÁRIO

1. Introdução	1
1.1 Contextualização	1
1.2 Motivação	3
1.3 Caracterização do Problema.....	7
1.4 Proposta de Solução	9
1.5 Estrutura da Dissertação	10
2. Metodologia da Pesquisa Científica	11
2.1 Metodologia	11
2.1.1 Método de Pesquisa	12
2.1.1.1 Estudo de Caso	13
2.1.1.2 Fases da Pesquisa Qualitativa	14
2.1.1.3 População e Amostra	15
3. Trabalhos Relacionados	17
3.1 Influência das opiniões no processo de decisão dos consumidores.....	17
3.2 Volume das opiniões e seus impactos nos consumidores	18
3.3 Confiabilidade e envolvimento dos consumidores nas redes sociais	19
3.4 Diversidade de tipos de consumidores e de fontes de informação	22
4. Fundamentação Teórica.....	24
4.1 Economia Compartilhada.....	24
4.2 Plataformas Ponto-a-Ponto de Consumo e Compartilhamento Colaborativo	26
4.3 Diretório Discover2Share	27
4.4 Análise de Sentimentos ou Mineração de Opiniões	32
4.4.1 Níveis de Análise.....	34
4.4.2 Abordagens de Análise de Sentimento	35
4.4.2.1 Abordagem baseada em Dicionário	35
4.4.2.2 Abordagem baseada em Aprendizado de Máquina.....	35
4.4.2.3 Abordagem Estatística	36
4.4.2.4 Abordagem Semântica	37
4.4.3 Análise de Sentimentos Multi-Idioma	37
4.4.4 Apresentação dos Resultados da Mineração de Opiniões.....	38
4.4.4.1 Apresentação a nível de documento.....	39
4.4.4.2 Apresentação a nível de característica	39
4.4.4.3 Abordagem híbrida	40
4.4.5 Ferramentas para Análise de Sentimentos - iFeel.....	41
4.4.5.1 Affin.....	41
4.4.5.2 Emolex.....	42
4.4.5.3 Emoticons	43
4.4.5.4 EmoticonDS.....	43
4.4.5.5 NRC Hashtag Sentiment Lexicon	43
4.4.5.6 Opinion Lexicon	44
4.4.5.7 Sentiment140 Lexicon	44

4.4.5.8	Umigon	45
4.4.5.9	VADER.....	46
4.4.5.10	PANAS-t.....	47
4.4.5.11	Happiness Index.....	48
4.4.5.12	SenticNet.....	48
4.4.5.13	SentiStrength.....	48
4.5	Processos Intensivos em Conhecimento	49
5.	Método da Proposta de Solução	52
5.1	Fase 1: Coleta dos Dados	52
5.2	Fase 2: Pré-processamento.....	54
5.3	Fase 3: Análise dos Dados	55
5.4	Fase 4: Apresentação	58
6.	Aplicação dos Estudos de Caso.....	59
6.1	Aplicação em um caso real do Método da Proposta de Solução	59
6.1.1	Aquisição dos Dados	60
6.1.2	Pré-processamento.....	60
6.1.2.1	Avaliação do Método Multi-Idioma	61
6.1.3	Apresentação dos Resultados	63
6.1.3.1	Visão Geral dos Dados Disponíveis.....	64
6.1.3.2	Frequência das Polaridades.....	65
6.1.3.3	Variação das Polaridades em uma Frequência Temporal	65
6.1.3.4	Nuvem de Termos mais comentados	67
6.1.3.5	Grid de Opiniões e Polaridades.....	68
6.2	Aplicação do Questionário On-line.....	70
6.2.1	Análise do Resultado da Aplicação do Questionário On-line.....	71
7.	Conclusão	77
7.1	Contribuições	78
7.2	Trabalhos Futuros	79
	Referências Bibliográficas	80

ÍNDICE DE FIGURAS

Figura 1	Informações da Airbnb no diretório Mesh	04
Figura 2	Informações da HomeAway no diretório Mesh	04
Figura 3	9Flats, Airbnb e HomeAway no Collaborative Consumption	05
Figura 4	Ontologia do Discover2Share	30
Figura 5	Radial contendo o Triângulo de Opiniões sobre Hotéis	40
Figura 6	Processo Intensivo em Conhecimento – Discover2Share	51
Figura 7	Visão geral do Método da Proposta de Solução	52
Figura 8	Coleta de Dados	53
Figura 9	Pré-processamento dos Dados	55
Figura 10	Análise de Sentimentos utilizando a ferramenta iFeel	57
Figura 11	Total de Dados coletados por cada plataforma	64
Figura 12	Qt. de opiniões por cada plataforma em valores absolutos	65
Figura 13	Qt. de opiniões por cada plataforma em valores percentuais	65
Figura 14	Frequência Temporal da Plataforma Airbnb	66
Figura 15	Frequência Temporal da Plataforma 9Flats	66
Figura 16	Frequência Temporal da Plataforma HomeAway	67
Figura 17	Nuvem de Termos mais mencionados	68
Figura 18	Grid de visualização de Opiniões, datas, e polaridades	69
Figura 19	Quantidade de Respostas obtidas no Estudo de Caso	71
Figura 20	Menções a atributos do Discover2Share	72
Figura 21	Mudança na Plataforma escolhida	73
Figura 22	Fatores citados como relevantes para a tomada de decisão	74
Figura 23	Qt. de respostas obtidas sobre a validade do estudo de caso	75

ÍNDICE DE TABELAS

Tabela 1	Comparativo de Plataformas no Discover2Share	06
Tabela 2	Fatores e Hipóteses de Pesquisa	20
Tabela 3	Conceitos e Fundamentação da Ontologia base do Discover2Share	29
Tabela 4	Conjuntos de Dados Relevantes	31
Tabela 5	Termo com sentimento negativo contido nos léxicos do Umigon	46
Tabela 6	Termo com sentimento positivo contido nos léxicos do Umigon	46
Tabela 7	Termos contidos na escala PANAS-x	47
Tabela 8	Visão Geral dos Dados Coletados em dois meses	60
Tabela 9	Percentual de Acertos em relação aos dados totais	62
Tabela 10	Percentual de Acertos em relação aos dados com polaridade	62

1. Introdução

Nesse Capítulo é apresentada a introdução dessa dissertação de mestrado e que está assim organizada: na Seção 1.1 esse trabalho é contextualizado. Na Seção 1.2 está contida sua motivação. A seguir, na Seção 1.3, é apresentada a caracterização do problema a ser tratado. Na Seção 1.4 é apresentada a proposta de solução e, por fim, na Seção 1.5, é apresentada a estrutura dessa dissertação.

1.1 Contextualização

As relações de consumo mudaram ao longo dos últimos anos, fazendo crescer a preocupação com seu impacto no meio ambiente, na sociedade em geral e no desenvolvimento. Ao mesmo tempo, crescem problemas como consumo elevado, a poluição e a pobreza. Em contrapartida, emerge também a discussão em como enfrentar tais problemas. Nesse sentido, tendo como base o papel fundamental dos indivíduos, ganha força um novo modelo de negócios, a Economia Compartilhada.

Essa nova tendência, cujo foco principal é o compartilhamento, é considerada um sistema socioeconômico construído ao redor do compartilhamento de recursos humanos e físicos, onde estão incluídas a criação, a produção, a distribuição, o comércio e o consumo compartilhado de bens e serviços por pessoas e organizações [1]. Tais iniciativas comerciais englobam transações como o compartilhamento, aluguel, empréstimo, doação, troca e o escambo [2]. Outra forma de compreender a economia compartilhada é vê-la como um conjunto de iniciativas de consumo conectado, onde é

ênfatisado o reuso de produtos e as conexões ponto-a-ponto (P2P) – ou seja, a comunicação feita diretamente entre os consumidores - eliminando-se assim os intermediários e possibilitando as interações face a face [3].

Ainda neste âmbito cabe destacar as Plataformas Ponto-a-Ponto (P2P) de Compartilhamento e Consumo Colaborativo (SCC) e o seu importante papel dentro da Economia Compartilhada ao permitirem que os indivíduos compartilhem recursos entre si.

Sobre as plataformas, Hoffen et al. [4] pontua que, à medida que o número dessas plataformas cresce, este novo ambiente de mercado se torna fragmentado, uma vez que as plataformas tentam se destacar e se diferenciar não só pelo tipo de recurso compartilhado ou colaborativamente consumido, mas também se dirigindo a diferentes regiões do mundo, às diferentes motivações de consumo e às diferentes formas de mediação, entre outras estratégias.

Do ponto de vista dos usuários dessas plataformas, tal fragmentação, além de tornar dispendioso, aumenta o esforço em encontrar uma plataforma adequada para oferecer ou acessar um determinado recurso. Para reforçar tal afirmação, tomemos alguns exemplos de plataformas que fornecem recursos de transportes, mais precisamente carros ou de acomodações. No primeiro caso temos opções como Uber, Karzoo, Amovens, Ride Joy, Unicaronas, entre outros. Para o segundo, a gama de opções é ainda maior. Temos, por exemplo: Airbnb, 9Flats, HomeAway, InterHome, 3rd Home, Vacation Home Rentals, entre tantas outras. Uma análise mais aprofundada nessas plataformas mostra ainda que há opções globais, atuando na maioria dos países, e outras que atuam apenas localmente, em regiões ou cidades específicas.

1.2 Motivação

A grande quantidade e também a diversidade das plataformas existentes, assim como a importância que essas têm ganhado no cotidiano dos indivíduos, gerou a necessidade de iniciativas que facilitem tanto o processo de oferta quanto o de pesquisa de recursos pelos usuários. Entre tais iniciativas destacam-se os websites de busca e descoberta de plataformas, com ênfase para o Mesh¹, o Collaborative Consumption² e o Discover2Share³.

O Diretório Mesh é uma iniciativa proposta por Gansky [1]. Trata-se de um diretório, atualizado pelos próprios usuários, onde é possível cadastrar plataformas enredadas, em diferentes categorias – finanças, moda, imobiliário, alimentação, etc. No momento da escrita desse trabalho havia 9.856 empreendimentos cadastrados no Diretório, distribuídos em 25 diferentes categorias e presentes em 132 países.

¹ Diretório Mesh: <http://meshing.it>

² Diretório Collaborative Consumption: <http://www.collaborativeconsumption.com>

³ Diretório Discover2Share: <http://www.discover2share.net>

travel > Airbnb

Airbnb

www.airbnb.com



Airbnb is the online marketplace allowing anyone, from private residents to commercial property managers, to rent out their extra space.

connect

share

Like 5

G+

Pin it

videos



Airbnb

recent activity

category



travel



real estate

contact



info@airbnb.com

location

99 Rhode Island Street, San Francisco, CA 94103, USA



service region

international

funded by

Sequoia, Y combinator, Youniversity, Greylock, Andreasson Horowitz, General Catalyst, Crunchfund, Founders Fund

Figura 1 – Informações da Airbnb no diretório Mesh [Elaboração própria]

travel > HomeAway

HomeAway

www.homeaway.com/



HomeAway offers an extensive selection of vacation homes that provide travelers with memorable experiences and benefits, especially more room to relax, for less than the cost of traditional hotel accommodations.

connect

share

Like 0

G+

Pin it

recent activity

category



travel

contact



<http://support.homeaway.com/>

location

1011 W. Fifth Street Suite 300, Austin 78703, TX, united_states

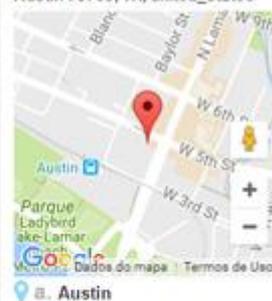


Figura 2 - Informações da HomeAway no diretório Mesh [Elaboração própria]

Nas figuras 1 e 2 é possível visualizar as informações de duas plataformas (Airbnb e HomeAway) no diretório Mesh. Como pode ser notado, esse diretório exibe algumas poucas informações sobre as plataformas: são fornecidas pequenas descrições, o endereço e também alguns meios de contato direto com a plataforma.

O Collaborative Consumption é um recurso online focado no consumo colaborativo ao redor do mundo, além de uma rede para a comunidade global. Nele são reunidos notícias, conteúdo, eventos e estudos sobre tal assunto. Além disso, o mesmo conta com um Diretório com uma base de dados composta por companhias de qualquer setor e serviço. No momento da escrita desse trabalho havia 16 categorias disponíveis e 1423 empreendimentos cadastrados.

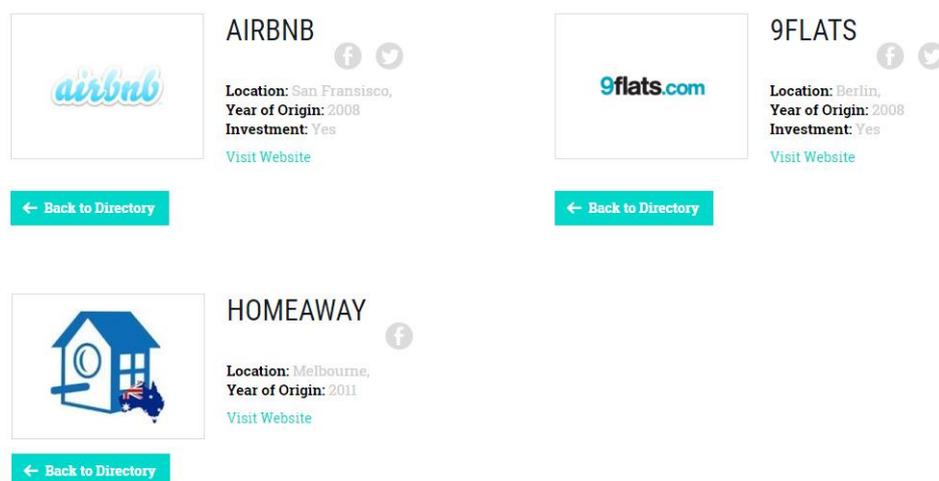


Figura 3 - Informações da 9Flats, Airbnb e HomeAway no diretório Collaborative Consumption [Elaboração própria]

Na Figura 3 são exibidas as informações existentes no diretório Collaborative Consumption sobre as plataformas Airbnb, 9Flats e HomeAway. De forma semelhante ao que acontece no diretório Mesh, são fornecidos pouquíssimos dados sobre as plataformas, como o local de origem, ano de fundação e links para os sites oficiais.

O Diretório Discover2Share é uma iniciativa mais recente, proposta na Universidade de Münster, Alemanha, e encabeçada por Hoffen et al [4]. Nesse Diretório é possível navegar pelos tipos de recurso/categorias disponíveis ou pelas regiões que contém recursos compartilhados. Um importante diferencial dessa proposta é a utilização de uma Ontologia como linguagem de descrição, o que permite a inserção estruturada e organizada de informações (Tabela 1) a respeito das plataformas compartilhadas. Esse Diretório será melhor detalhado na Seção 4.3.

Tabela 1. Comparativo de Plataformas baseado em dados do Discover2Share

	Airbnb	9Flats	HomeAway
Type	Platform; Named Individual	Platform; Named Individual	Platform; Named Individual
Foundation Place	San Francisco – CA - USA	Hamburg – DEU	Austin – TX - USA
Residence	San Francisco – CA - USA	Hamburg – DEU	Austin – TX - USA
Accepted Payment Methods	Per Transaction	Per Transaction	Per Transaction; Membership-fee; Combination
Accept Physical Resources Of	Accommodations	Accommodations	Accommodations
Consumer Involvement	In Between	In Between	In Between
Duration Of Service: Min / Max	Days / Days	Days / Weeks	Days / Weeks
Grants To Be Resource Provider	Private	Private	Private and Business
Has Access Type	Integrated	Integrated	Integrated
Has Mobile Application Support	Android; IOS; Windows	--	Android; IOS
Money Flow	C2B2C	C2B2C	C2B2C
Motives For Consumption	Economical; Social	Economical; Social	Economical; Social
Necessary Time For Transaction	Deferred	Deferred	Deferred
Profit Structure	Profit From Both Peer Consumers And Peer Providers	Profit From Peer Providers	Profit From Peer Providers
Served Area	Global	Global	Global
Type Of Accessed Object	Mixed	Mixed	Mixed
Founding Year	2008	2011	2005

Id Number	190	777	258
Title	Airbnb	9Flats	HomeAway
URL	http://www.airbnb.com	http://www.9flats.com	http://www.homeaway.com
Identifier	190	777	258

1.3 Caracterização do Problema

Tomando como exemplo o Diretório Discover2Share, ao analisar os processos que compreendem a busca, navegação e escolha de plataformas realizadas pelos seus usuários, vemos caracterizado um Processo Intensivo em Conhecimento (ou PIC, proveniente do inglês KIP — Knowledge Intensive Process) [56], uma vez que nos mesmos estão envolvidos conhecimentos tácitos. Além disso, alguns passos demandam a tomada de decisão e a análise dos resultados obtidos antes que o usuário siga para os próximos passos.

Ao final desses processos intensivos em conhecimento, são apresentadas aos usuários as plataformas que atendem aos critérios utilizados. Tais informações consistem em dados estruturados, conforme pode ser visto na Tabela 1, onde são apresentados os dados recuperados do Diretório relativos às plataformas (nesse caso, 9Flats, Airbnb e HomeAway).

O Discover2Share apresenta alguns importantes diferenciais frente aos demais diretórios. Por exemplo, a comparação pelos atributos de cada plataforma (Tabela 1) não está disponível nas demais iniciativas. Nelas, normalmente, constam algumas poucas e diferentes informações a respeito das plataformas. Em alguns casos são disponibilizados apenas o nome e o link para a página oficial.

Com base na quantidade e qualidade dos dados disponibilizados nos diretórios apresentados, o Discover2Share se destaca como a melhor opção como canal para a

busca de plataformas. Entretanto, ser o melhor diretório entre os diretórios comparados não garante, por si só, possibilitar aos seus usuários uma escolha mais adequada.

Para embasar a constatação acima, pode-se imaginar um cenário onde um usuário, que deseja viajar, por exemplo, acesse o Discover2Share tendo em mente encontrar a plataforma mais apropriada para lhe fornecer tal recurso. Ao se deparar com as informações hoje existentes, e exibidas na Tabela 1, ele descobriria quando as plataformas foram criadas, qual o meio de pagamento que aceitam, a duração mínima e máxima do serviço, etc. Embora tratem-se de informações que podem ajudar a responder algumas de suas dúvidas, elas dificilmente serão suficientes para permitir a escolha mais adequada. Comparando as três plataformas já mencionadas, cujos dados são bastante parecidos, o usuário poderia, então, optar pela mais antiga delas, imaginando que existir há mais tempo faz com que a plataforma seja a mais conhecida ou a mais estável. Nesse caso, a plataforma escolhida seria a HomeAway. Entretanto, uma rápida análise nos dados oriundos das opiniões de outros usuários, que serão apresentadas e aprofundadas mais adiante, mostra que a plataforma mais conhecida, mais comentada e também a mais utilizada é a Airbnb.

Logo, conclui-se que, embora úteis, as informações hoje existentes no Discover2Share tendem a não ser suficientes para permitir escolhas mais assertivas.

Considerando a importância da economia compartilhada e, sobretudo, das plataformas ponto-a-ponto de consumo e compartilhamento colaborativo, assim como do Discover2Share como canal para busca e descoberta de plataformas, têm-se definido o problema a ser tratado nesse trabalho:

“Como enriquecer o Diretório Discover2Share a fim de melhorar o processo de identificação de plataformas ponto-a-ponto de consumo e compartilhamento colaborativo?”

1.4 Proposta de Solução

Naturalmente, quando diante de situações onde precisa diminuir as dúvidas frente à escolha dentre diferentes alternativas, os indivíduos tendem a confiar nas recomendações de outros indivíduos, sejam essas obtidas de forma direta, seja por meio de textos de recomendação, opiniões ou reviews [5]. Além disso, quando uma pessoa se vê diante de uma situação onde se sente incapaz de avaliar adequadamente as características de um produto através da observação direta e do contato, ela busca as reações dos outros como evidência sobre a verdadeira natureza do produto [6].

Tal fator social é ainda mais potencializado pela quantidade de informações disponíveis on-line, onde o acesso dos indivíduos à informações sobre produtos e serviços é muito mais eficiente e flexível. Conseqüentemente, os *reviews* publicados por outros usuários afetam consideravelmente a qualidade de suas decisões e avaliações.

Frente à necessidade de enriquecer as informações hoje existentes no Discover2Share e melhorar o processo de identificação da plataforma mais adequada, a solução ora proposta fará uso justamente das opiniões de usuários e consumidores em geral disponíveis on-line. Tal conteúdo, normalmente denominado de conteúdo social, está amplamente disponível em blogs, sites e, sobretudo, nas redes sociais. No âmbito desse trabalho serão utilizados os textos provenientes das redes sociais Twitter e Facebook.

Defende-se nesse trabalho que:

“O uso das opiniões provenientes de conteúdo social enriquecerá, de forma semântica, os dados relativos às plataformas existentes atualmente no Diretório Discover2Share, melhorando assim o processo de identificação de plataformas para se utilizar pelos seus usuários.”

1.5 Estrutura da Dissertação

O restante dessa dissertação está estruturado da seguinte forma: no Capítulo 2 é apresentada a Metodologia da Pesquisa Científica. A seguir, no Capítulo 3, são apresentados os Trabalhos Relacionados, onde serão vistos trabalhos que tratam da influência, do impacto e da confiabilidade, entre outros fatores, das opiniões no processo de decisão dos consumidores. No Capítulo 4 é apresentada a Fundamentação Teórica que fornece o embasamento a esse trabalho. Neste serão vistos os conceitos relativos à Economia Compartilhada, às Plataformas Ponto-a-ponto de Consumo e Compartilhamento Colaborativo, à Análise de Sentimentos/Mineração de Opiniões e aos Processos Intensivos em Conhecimento. Também será apresentado o Diretório Discover2Share. No Capítulo 5 é exposto o Método da Proposta de Solução, sendo procedido pelo Capítulo 6, onde é apresentada a Aplicação dos Estudos de Caso realizados para avaliar a Proposta de Solução. Por fim, no Capítulo 7, é apresentada a Conclusão da presente dissertação.

2. Metodologia da Pesquisa Científica

Nesse Capítulo é apresentada uma breve discussão sobre a metodologia de pesquisa empregada nesse trabalho, a fim de fundamentar a escolha do método utilizado.

2.1 Metodologia

Uma pesquisa pode ser definida como um processo sistemático e formal no qual o método científico é desenvolvido de forma a permitir a descoberta de respostas para problemas através do uso de procedimentos científicos.

Segundo Matta [57] as pesquisas podem ser classificadas através de diversos critérios:

- **Natureza das variáveis:** nesse sentido, a pesquisa pode ser quantitativa ou qualitativa. No primeiro as hipóteses e variáveis são claramente definidas e são usadas para se obter uma medição precisa dos resultados quantificáveis obtidos. Na segunda se procura verificar um fenômeno através da observação e estudo do mesmo [58].
- **Objetivos e características:** nesse caso as pesquisas são classificadas em exploratória, descritiva ou casual. A exploratória é usada quando não se tem muito conhecimento sobre determinado assunto. Normalmente utilizada quando se pretende obter um primeiro contato com a situação a ser pesquisada, sendo seu principal objetivo a descoberta. A descritiva é empregada quando já se possui algum conhecimento a respeito do assunto e se pretende descrever um fenômeno. Logo, algumas hipóteses podem ser formuladas tendo como base prévios conhecimentos, e procura-se então

confirmá-las ou negá-las. A causal é normalmente empregada quando se pretende provar que uma determinada variável exerce influência sobre outras, numa relação de causa-efeito. Seu objetivo é juntar evidências e descobrir a relação existente.

- **Escopo:** nesse sentido a pesquisa pode ser classificada como Estudo de Caso, Levantamento Amostral e Estudo de Campo. O primeiro é um estudo profundo e não amplo, onde se procura conhecer a fundo apenas um ou poucos elementos da população a respeito de um grande número de aspectos e suas inter-relações. O Levantamento Amostral se preocupa com a obtenção de dados que tenham representatividade na população estudada, tanto no que diz respeito ao número quanto do processo de seleção dos elementos da amostra da pesquisa. Já o Estudo de Campo se situa no meio dos dois anteriores, trabalhando com amostras de dimensões que possibilitem análises estatísticas sem, entretanto, ter preocupações com a representatividade da amostra.

Tendo como base a fundamentação acima, esse trabalho pode ser classificado como uma pesquisa qualitativa exploratória que fará uso do estudo de caso como pesquisa de campo. A seguir cada um desses critérios será melhor detalhado.

2.1.1 Método de Pesquisa

O objetivo definido na Proposta de Solução e defendido nesse trabalho pode ser dividido em dois:

1. **Agregação de mais dados:** o uso das opiniões provenientes de conteúdo social enriquecerá os dados existentes atualmente, relativos às plataformas, no Diretório Discover2Share; e

2. **Melhora do processo de identificação:** as opiniões provenientes de conteúdo social fornecerão um contexto aprimorado para que os usuários possam melhor identificar uma plataforma para utilizar.

Como mencionado anteriormente, a utilização de opiniões no ambiente online como fator impactante no processo de tomada de decisão é um tema relativamente novo. Logo, para avaliar cada objetivo descrito acima foi utilizado um Método Qualitativo com caráter exploratório e que faz uso de um Estudo de Caso aplicado a um grupo de usuários. As fases da pesquisa e o seu resultado serão discutidos a seguir.

2.1.1.1 Estudo de Caso

Por oferecer uma forma de entendimento mais global e compreensiva do fenômeno que se deseja estudar, permitindo ainda trazer à tona considerações contextuais relevantes e que normalmente não são reveladas por outros métodos [59], o Estudo de Caso foi o método escolhido para ser aplicado nesse trabalho.

Conforme Yin [60], o Estudo de Caso é indicado para tratar de questões sobre eventos atuais, a respeito dos quais o pesquisador tem limitado ou nenhum controle. Mesmo que não produza conclusões generalizáveis para toda a população, permite que seja verificada a adequação de conceitos, expandindo assim e confirmando teorias que podem vir a servir de referência em estudos futuros. Ainda segundo Yin, existem três características que delineiam o Estudo de Caso:

- Obtenção dos dados em um nível de profundidade que permita ao mesmo tempo caracterizar e explicar detalhadamente aspectos singulares do estudo e também apontar semelhanças e diferenças quando comparado com outros casos estudados;
- O pesquisador se preocupa com a busca de conhecimento e não por conclusões e verificações;

- O pesquisador deve ser capaz de reunir e interpretar vários aspectos do objeto pesquisado.

2.1.1.2 Fases da Pesquisa Qualitativa

A realização da pesquisa compreendeu a realização dois Estudos de Caso: um através da aplicação em um caso real do método proposto (descrito em detalhes no Capítulo 5) e o outro através de um questionário on-line.

Em relação ao questionário, tal instrumento foi assim estruturado:

- **Introdução/Motivação:** foi utilizado um pequeno texto de introdução com o objetivo de explicar o objetivo do Estudo de Caso. Além disso, tal texto contextualizou os dois conjuntos de dados, separados cada um em uma página (ver Anexos 1 e 2), contidos no estudo;
- **Conjuntos de dados:** o questionário foi composto por dois conjuntos de dados. O primeiro contendo os dados existentes hoje no Discover2Share e dispostos em uma tabela e o segundo contendo os dados obtidos na avaliação da abordagem e apresentados através dos elementos visuais detalhados no Capítulo 5.
- **Questões do primeiro conjunto de dados:** ao final da primeira página, após ser apresentada a tabela contendo o primeiro conjunto de dados, foram inseridas duas perguntas. A primeira, de tipo fechado, objetivava saber do usuário/participante qual plataforma ele escolheria levando em consideração o conjunto de dados apresentado. A resposta a essa questão era mutuamente exclusiva, ou seja, só foi possível escolher uma das três opções (9Flats, Airbnb ou HomeAway). A segunda pergunta, de tipo aberto, perguntava o participante qual o motivo da escolha realizada na primeira questão.

- **Questões do segundo conjunto de dados:** para visualizar o segundo conjunto de dados o participante, obrigatoriamente, precisou responder às questões do primeiro conjunto. Nessa segunda página, após ser apresentado o segundo conjunto de dados, foram inseridas três questões. A primeira, de tipo fechado, tinha como objetivo saber do participante qual plataforma ele escolheria tendo como base o segundo conjunto de dados apresentado. Tal questão foi semelhante à primeira do conjunto anterior e, portanto, também era mutuamente exclusiva. A segunda questão, de tipo aberto, perguntava o participante qual o motivo da escolha realizada e a última, também de tipo aberto, perguntava ao participante qual a opinião dela a respeito do experimento em questão.

2.1.1.3 População e Amostra

Na investigação de um problema, a população corresponde ao grupo de todos os elementos que compartilham um conjunto comum de características de interesse [61]. Nesse sentido, a população alvo desse estudo pôde ser definida por todos os indivíduos que já tenham utilizado uma plataforma P2P de consumo e compartilhamento colaborativo, pelos que tem interesse em utilizar e, em último caso, pelos que tem alguma familiaridade com o tema em questão.

A respeito das amostras, os estudos quantitativos dependem de grandes amostras selecionadas de forma aleatória e que possibilitem inferências e generalizações em relação a toda uma população. Por outro lado, os estudos qualitativos têm como preocupação o estudo em profundidade, fazendo uso de amostras pequenas e até mesmo de casos únicos intencionalmente selecionados [62].

Com relação ao processo de seleção de amostras, Babbie [63] apresenta algumas técnicas:

- **Amostragem probabilística:** nesse caso, todos os elementos da população possuem probabilidade conhecida, diferente de zero, de serem selecionados para fazerem parte da amostra.
- **Amostragem não probabilística:** aqui, as amostras são selecionadas por critérios subjetivos, de acordo com a experiência do pesquisador ou com os fins do estudo. Logo, não há uma probabilidade conhecida de um indivíduo da população ser selecionado. A amostragem probabilística pode ser dividida em alguns tipos:
 - **Por conveniência:** as amostras são selecionadas conforme a conveniência do pesquisador, sendo constituídas de indivíduos que estão ao seu alcance e que se dispuseram a responder ao questionário.
 - **Por julgamento:** a seleção aqui ocorre conforme um critério de julgamento, onde o pesquisador escolhe a amostra com base no que acredita que ela possa contribuir para o estudo.
 - **Por cota:** são selecionadas amostras que se identifiquem, de alguma forma, com a população.

Neste trabalho optou-se por utilizar as técnicas não probabilistas na escolha das amostras. Com isso, foram selecionados elementos por conveniência, ou seja, indivíduos que estavam ao alcance e dispostos a participarem; por julgamento e por cota, sendo escolhidas pessoas dentro do critério de população-alvo acima delimitado.

Cabe pontuar ainda que, dada a sua natureza, não é possível, ao se fazer uso de amostras obtidas pelas técnicas não probabilísticas, inferir sobre um universo, uma vez que é desconhecido o erro cometido na escolha dos elementos que comporão a amostra.

3. Trabalhos Relacionados

Esse Capítulo apresenta os Trabalhos Relacionados, onde são vistos outros trabalhos que tratam da utilização de conteúdo social como insumo para a tomada de decisões.

3.1 Influência das opiniões no processo de decisão dos consumidores

A influência de outras opiniões no processo de tomada de decisão de compra dos consumidores tem sido discutida e comprovada em uma série de pesquisas ao longo dos anos [7; 8; 9; 10; 11]. Da mesma forma, outras pesquisas têm tratado da crescente necessidade demonstrada pelos consumidores on-line na interação social como contribuição adicional para suas metas de consumo [12; 13; 14; 15].

O processo de tomada de decisão no ambiente on-line é fortemente impactado por todo o engajamento possibilitado por canais como fóruns, blogs, ferramentas de avaliação e, sobretudo, pelas redes sociais que fornecem os meios necessários para que os consumidores busquem informações sobre o produto ou serviço que desejam adquirir. Embora vários pesquisadores defendam essa tese, cabe citar o trabalho de outros autores que vão de encontro à mesma questionando, sobretudo, a credibilidade e a qualidade das informações online. Tais pesquisadores defendem que trocar ou compartilhar informações com pessoas desconhecidas por meio da internet é falível e não confiável. Dessa forma, segundo os mesmos, é menos provável que as opiniões obtidas no meio online tenham impacto na tomada de decisão do consumidor. Algumas referências dessa corrente podem ser obtidas em [16].

Esse trabalho será orientado tendo em vista a primeira corrente, frente às duas acima citadas, ou seja, a validade do uso de opiniões como conteúdo rico e relevante na

tomada de decisão. Nesse sentido, a seguir serão discutidos alguns trabalhos recentes que tratam justamente desse tema.

3.2 Volume das opiniões e seus impactos nos consumidores

Na Seção anterior foram apresentadas as correntes distintas que tratam da influência das opiniões no processo de decisão dos consumidores. Nessa Seção, considerando que a linha adotada nesse trabalho é a que toma as opiniões de outros consumidores como importante insumo no processo de tomada de decisão, será abordado o grande volume existente de opiniões e o seu impacto nos consumidores.

Em [17], inicialmente é discutida a utilidade das opiniões. Em tal pesquisa é proposto um modelo, baseado em aprendizado de máquina, que pode prever a utilidade dos *reviews* dos consumidores utilizando vários recursos textuais como polaridade, subjetividade, entropia e facilidade de leitura.

O contexto apresentado nesse trabalho trata do grande volume de opiniões existentes atualmente e do impacto positivo e negativo que esses milhares de dados não estruturados causam nos consumidores. Tratando mais especificamente das avaliações disponíveis, é apresentado como exemplo o site da Amazon [www.amazon.com], onde é praticamente impossível que os consumidores consigam ler todos os *reviews* disponíveis, uma vez que alguns produtos, menos populares, podem ter centenas de avaliações e os populares milhares de opiniões postadas. Em tais situações, onde grandes volumes de dados estão disponíveis para serem analisados, sua utilidade para os consumidores poderia ser incrementada se houvesse um maior nível de visibilidade das críticas realmente úteis e que reflitam uma visão geral do produto ou serviço. Para tal, mecanismos simples como perguntar aos próprios usuários se um determinado comentário foi útil ou não ajudam a criar um sistema de ranqueamento que considera a

utilidade do comentário. Tal sistema de ranqueamento, baseado no mencionado conceito de utilidade do *review*, é o cerne de [17].

Partindo da forma usual de atribuição de utilidade a um comentário, ou seja, feito manualmente pelos próprios usuários, o modelo proposto por Singh et al. [17] atribui automaticamente valores de utilidade a um *review* logo que o mesmo é postado, a fim de que ele tenha uma chance justa de ser visualizado por outros compradores. Além disso, outras contribuições foram previstas por esse trabalho, como ajudar aos compradores a escreverem melhores críticas e, conseqüentemente, ajudar aos novos compradores a tomarem sua decisão de compra, além de auxiliar as empresas a melhorarem seus sites.

Esse trabalho trouxe como contribuição para essa dissertação a abordagem utilizada frente a uma situação de informação intensiva, onde um sistema de ranqueamento automatizado foi utilizado a fim de melhorar a qualidade da informação, mais precisamente dos reviews postados por outros usuários, disponível para os consumidores on-line.

3.3 Confiabilidade e envolvimento dos consumidores nas redes sociais

Em [18] é abordado o boca-a-boca eletrônico — *eletronic word-of-mouth* ou simplesmente ‘e-WOM’, conceito que pode ser definido como a comunicação entre consumidores sobre um produto, serviço ou companhia em que as fontes são consideradas independentes de influência comercial. Nesse trabalho, o ‘e-WOM’ é observado em um domínio específico, o turismo, através do uso do Twitter pelos turistas e do seu impacto na área em questão. A proposta dessa pesquisa foi a criação de um framework conceitual para compreender os fundamentos da comunicação digital e

investigar, empiricamente, sua validade examinando os fatores que influenciam o comportamento do consumidor no turismo.

No desenvolvimento do artigo foi apresentada a fundamentação da importância das Redes Sociais para os destinos turísticos, assim como para os negócios envolvidos, com destaque para a mudança provocada pelas redes sociais no que diz respeito à forma como os turistas buscam, leem, confiam e também produzem informações colaborativas sobre fornecedores de turismo e destinos turísticos. Dessa forma, ao coproduzirem e compartilharem uma grande quantidade de informações e conhecimento – o conhecido “conteúdo gerado por usuários” – eles se tornam co-comerciantes, co-designers, co-produtores e co-consumidores de experiências de viagens e turismo.

Com o intuito de avaliar o framework criado e também testar a validade das hipóteses propostas (tabela 2), foi utilizado um método quantitativo, através da aplicação de um questionário online contendo 48 questões.

Tabela 2. Fatores e Hipóteses de Pesquisa [18]

Fator	Hipóteses
Confiabilidade/credibilidade da fonte	H1-1: usuários do Twitter, acreditando que outros usuários/seguidores/grupos são confiáveis, tendem a usar informações obtidas do Twitter sobre serviços turísticos. H1-2: usuários do Twitter, estimando que outros usuários/seguidores/grupos não são confiáveis, tendem para usar informações obtidas do Twitter sobre serviços de turismo.
Grau de envolvimento	H2-1: usuários do Twitter, acreditando que a compra de serviços de turismo constitui um processo de decisão de alto envolvimento, tendem a gastar mais tempo na postagem. H2-2: usuários do Twitter que estimam que a compra de serviços de turismo constitui um processo de decisão de alto envolvimento tendem a usar informações relativas fornecidas por outros usuários/seguidores.
Frequência de comunicação	H3: usuários do Twitter frequentemente se comunicam com seus seguidores e amigos, tendendo a usar informações obtidas do Twitter sobre serviços turísticos.
Experiência e conhecimento das fontes	H4: usuários do Twitter estimam que outros usuários/seguidores podem ter experiência e/ou conhecimento sobre serviços turísticos específicos e tendem a usar essa informação.

Ao final, a partir da análise do conjunto de respostas obtido na aplicação do questionário, as hipóteses foram validadas, conforme segue:

- A primeira hipótese, a confiabilidade de outros usuários e seguidores, foi confirmada. Consequentemente, também foi confirmada a teoria de que fontes confiáveis tem um papel influente no uso da informação turística.
- Quanto à segunda hipótese, relacionada ao nível de envolvimento, os resultados indicaram que um alto grau de envolvimento no processo de compra de serviços turísticos leva a um alto grau de postagem. Por outro lado, o grau de envolvimento não leva os usuários a usarem informações do Twitter para tomarem sua decisão. Nesse caso, eles provavelmente utilizam mídias adicionais como fonte de informação no processo de decisão de compra.
- A terceira hipótese era destinada a explorar a questão da comunicação e da confiança relativa. Nas redes sociais existe uma ligação/relação forte ou fraca com os seguidores ou amigos dependendo da frequência de comunicação. A literatura relacionada sugere que se a conexão for considerada forte e crível, o receptor do tweet/mensagem usa as informações fornecidas. Os resultados obtidos mostraram que essa hipótese não está confirmada; as ligações de grupo/pessoa influenciam o usuário, mas em um grau relativamente baixo.
- A última hipótese explorava a questão da credibilidade da fonte a partir de uma perspectiva especializada. Quando um usuário/seguidor possui conhecimentos e expertise específicos, então as informações fornecidas pelo boca-a-boca eletrônico e pelos *reviews* online podem ser considerados mais credíveis e relevantes do que outras fontes de informação. Tal hipótese foi confirmada. Logo, os especialistas têm um papel influente e os usuários aproveitam seus conhecimentos e expertise.

Resumidamente, os resultados do estudo em questão, onde o framework sugerido foi empiricamente testado através de um método qualitativo focado na perspectiva dos recebedores das informações/tweets (e não na perspectiva dos narradores/remetentes das informações), indicaram que três fatores são muito influentes quanto ao uso de informações relativas a serviços turísticos obtidas do Twitter: confiabilidade dos seguidores/usuários do Twitter; nível de envolvimento – publicação; e expertise e conhecimento dos usuários/seguidores do Twitter. Por outro lado, os demais fatores investigados apresentaram um grau relativamente baixo de influência.

A importância desse trabalho para essa dissertação se dá sob algumas perspectivas, sendo a principal delas a confirmação, através de uma pesquisa científica fortemente fundamentada, realizada ao longo de cinco meses e que contou com a participação de cerca de 500 pessoas, da validade e utilidade do conteúdo social como importante insumo na tomada de decisão dos consumidores no ambiente online.

3.4 Diversidade de tipos de consumidores e de fontes de informação

Como visto ao longo desse Capítulo, os consumidores online são suscetíveis ao volume disponível de opiniões assim como à origem das mesmas, apresentando assim diferentes comportamentos em diferentes situações. Chen et al. [19] tratam, através da realização de um estudo de caso dividido em duas fases, dos efeitos das fontes de informação online na intenção de compra entre consumidores. Nesse trabalho, os consumidores foram separados em dois grupos: os que possuem alta susceptibilidade de influência por informações e os que possuem baixa susceptibilidade. Além disso, as fontes de informação online foram divididas em três grupos: boca-a-boca eletrônico; informações neutras ou produzidas por terceiros; informações dos próprios fabricantes ou vendedores. Ao longo desse trabalho os autores abordam questões sociais e

antropológicas a respeito das influências percebidas pelos indivíduos a partir de outras opiniões e pontos de vista.

O resultado do estudo de caso realizado pelos pesquisadores em questão reafirma que o processo de tomada de decisão dos consumidores é positivamente influenciado pelas opiniões provenientes do ambiente online. Nesse sentido, o boca-a-boca eletrônico é percebido como mais útil pelos consumidores com alta susceptibilidade do que as demais fontes tratadas. Em relação aos consumidores com baixa susceptibilidade, as três fontes tratadas são vistas como fontes úteis em sua tomada de decisão.

Assim como o trabalho descrito na Seção anterior, a contribuição desse último a esta dissertação se deu pela metodologia científica empregada – um estudo de caso – onde também foi comprovada a importância das opiniões para o processo de tomada de decisão dos consumidores.

4. Fundamentação Teórica

Ao longo deste trabalho várias áreas de pesquisa são abordadas, dando ao mesmo um caráter multidisciplinar. Conceitos como Economia Compartilhada e Plataformas de Consumo Compartilhado Colaborativo, Análise de Sentimentos e Mineração de Opiniões, uso de Conteúdo Social como insumo para a Tomada de Decisões e Processos Intensivos em Conhecimento são vistos e utilizados permeando tanto a definição do problema quanto, sobretudo, a proposta de solução. Neste Capítulo tais conceitos serão aprofundados, juntamente com a descrição do Diretório Web Discover2Share e da Ferramenta iFeel, fornecendo assim a fundamentação teórica que servirá de base para a dissertação como um todo.

4.1 Economia Compartilhada

Desde o seu início, a internet passou por muitos passos de evolução, gerando assim vários novos modelos de negócios eletrônicos, móveis e sociais. Recentemente, um novo paradigma tem ganhado força através do que tem sido chamado de “web social”, onde o foco deixa de ser o “possuir” e passa a ser o “usar” bens ou serviços. Tal fenômeno, que contraria o modelo tradicional que tem base na propriedade, tem sido chamado de Economia Compartilhada, onde o foco está justamente no compartilhamento de produtos e serviços, entre outros recursos. Botsman [20] e Hamari et al. [21] citam três das principais características desse novo modelo de negócios:

- **Mudança no comportamento de consumo:** enquanto a propriedade foi um modelo predominante para o uso de bens (por exemplo, carros) no passado, o uso temporário tornou-se, recentemente, mais atrativo para muitos

consumidores. Como exemplo de negócios com foco no uso temporário de bens, pode-se citar algumas companhias que oferecem acesso a serviços compartilhados de mobilidade através do aluguel de carros, como o Car2Go, ou bicicletas, como o Nextbike e o Green Bikes. Além do fator conveniência, alguns outros são normalmente citados para justificar o uso de tais serviços, como preços mais baixos e sustentabilidade ecológica.

- **Redes sociais e mercados eletrônicos:** a rede entre pares é mantida, principalmente, pelas redes sociais e pelas plataformas comunitárias. Tais locais fornecem o ambiente ideal para ligar consumidores dispostos a compartilharem bens entre si. Além disso, as plataformas eletrônicas de mercado reduzem consideravelmente o custo de pesquisa e de transação. Outros fatores importantes, vistos em tais plataformas, são os seus mecanismos de confiança e reputação, que costumam funcionar através de classificação e feedback, e também o atendimento integrado, que conta com funções de pagamento que asseguram uma compensação fácil e confiável para o uso de serviços compartilhados. Um exemplo desse tipo de mercado eletrônico é o Mercado Livre.
- **Dispositivos móveis e serviços eletrônicos:** os dispositivos móveis, como *smartphones* e *tablets*, tão comuns hoje em dia, são um grande facilitador para a “economia de aplicativos”. Além da conveniência de se compartilhar recursos fazendo uso de tais dispositivos, algumas companhias, como a já citada Car2Go e também a DriveNow, a Sharoo ou a RelayRides (todas fornecedoras de recursos de mobilidade) apostam ainda na combinação de serviços eletrônicos com o uso de aplicativos e cartões inteligentes no lugar de chaves físicas, por exemplo.

O termo “Economia Compartilhada” foi utilizado pela primeira vez em 2008. Desde então, tem sido utilizado para definir alguns conceitos, além de ganhar diferentes nomes, como Economia Mesh [1]; Consumo Colaborativo [2]; ou Consumo Conectado [3]. Em comum, independente do nome ou dos conceitos aos quais é associada, temos o conceito de que a Economia Compartilhada está relacionada ao consumo colaborativo feito através de atividades como o compartilhamento, o intercâmbio e o aluguel de recursos sem a posse dos bens em questão [22]. No contexto das transações econômicas, a Economia Compartilhada está relacionada ao uso de um recurso – bem físico ou serviço – cujo consumo é dividido em partes únicas. Tais partes são colaborativamente consumidas em redes C2C (*consumer-to-consumer* — modelo de negócios onde o comércio se desenvolve entre usuários pessoas físicas, também chamado “um para um”) coordenadas através de serviços baseados em comunidades on-line ou através de intermediários em modelos B2C (*business-to-consumer* — modelo de negócios onde o comércio é realizado entre empresas, sejam estas produtoras, vendedoras ou prestadores de serviço e o consumidor final) [22].

4.2 Plataformas Ponto-a-Ponto de Consumo e Compartilhamento Colaborativo

As plataformas ponto-a-ponto (P2P) de consumo e compartilhamento colaborativo (SCC) desempenham um importante papel dentro da Economia Compartilhada como facilitadoras entre as partes envolvidas no consumo compartilhado. Tendo normalmente como base os recursos tecnológicos e as mídias sociais, as plataformas P2P permitem uma conexão direta entre provedores e usuários de serviços e recursos sob demanda. A Airbnb - que permite o compartilhamento de recursos do tipo “acomodações” - e o Uber - que permite o compartilhamento de recursos de mobilidade – mais precisamente automóveis, são exemplos de atividades relacionadas à Economia Compartilhada e que

são suportadas por plataformas. Ambas fornecem tanto um website quanto um aplicativo que permitem aos seus usuários – tanto aquele que fornece o recurso quanto o que consome – estabelecerem relações de consumo entre si.

4.3 Diretório Discover2Share

Tendo em mente a diversidade e a grande quantidade de plataformas P2P SCC disponíveis atualmente, além da ausência de mecanismos que facilitem a sua descoberta de forma eficiente, Hoffen et al. [4] desenvolveram o Diretório Web Discover2Share.

Como ponto de partida, ao descreverem as motivações do trabalho em questão, os pesquisadores descrevem os principais problemas encontrados nos diretórios/catálogos web existentes atualmente que se propõem a realizar a tarefa de facilitar a descoberta de plataformas, a saber:

- Informações desencontradas: a meta-informação fornecida sobre as plataformas varia de catálogo para catálogo;
- Escassez de informações: normalmente é fornecido apenas um link para o site oficial da plataforma juntamente com uma pequena descrição sobre o serviço oferecido;
- Classificação heterogênea: as plataformas são classificadas com base em um conjunto heterogêneo de características, o que dificulta, por exemplo, a sua descrição e também a comparação com as demais existentes;
- Informações redundantes: os catálogos contêm informações redundantes sem meios efetivos para consolidar os dados devido às diferentes estruturas de dados;
- Mecanismos de controle: os catálogos não proveem mecanismos efetivos de controle para assegurar que as classificações das plataformas sejam consistentes e coerentes, em termos de sintaxe e semântica.

Por isso, Hoffen et al. propuseram o desenvolvimento do diretório web Discover2Share com o objetivo de facilitar a descoberta efetiva e transparente de plataformas ponto-a-ponto (P2P) de consumo e compartilhamento colaborativo (SCC — do inglês Sharing and Collaborative Consumption).

No processo de desenvolvimento do Discover2Share, outras contribuições foram alcançadas, como a criação de uma ontologia, cujo processo de desenvolvimento consistiu na pesquisa prévia de outras ontologias existentes a fim de possibilitar a reutilização de conceitos existentes e estabelecidos na literatura científica, como, por exemplo, o conceito de duração, proveniente da OWL-time⁴ (uma ontologia de conceitos temporais utilizada para descrever as propriedades temporais de recursos no mundo real ou descritos em páginas web) ou ofertas em *GoodRelations*⁵ (uma ontologia para e-commerce que permite a publicação pelas empresas de detalhes sobre seus produtos e serviços de forma amigável) e produtos e serviços em eClassOWL⁶ (foi desenvolvida para ser usada em combinação com a *GoodRelations*. Trata-se de uma ontologia utilizada para descrever os tipos e as propriedades de produtos e serviços na Web Semântica).

Na tabela 3 são apresentados alguns dos conceitos que compõem e fundamentam a ontologia criada com as suas respectivas referências na literatura. A lista completa pode ser encontrada em [4].

⁴ OWL-time: <https://www.w3.org/TR/owl-time/>

⁵ GoodRelations: <http://purl.org/goodrelations/>

⁶ eClassOWL: <http://www.heppnetz.de/projects/eclassowl/>

Tabela 3: Conceitos e Fundamentação da Ontologia base do Discover2Share [adaptado de 4]

Descrição dos Conceitos	Referências
Resource Type: descreve o tipo de recurso físico que um provedor (<i>peer-provider</i>) deve possuir ou ter acesso para poder prover o serviço a um consumidor (<i>peer-consumer</i>)	[23]
Type of Accessed Object: é um conceito que ajuda a esclarecer a natureza do acesso em um cenário P2P SCC, diferenciando entre o acesso a um objeto puramente funcional como um espaço de armazenamento e o acesso de experiência, como uma visita a um museu.	[24, 25]
Consumer Involvement: é usado para avaliar a medida em que o cliente está envolvido na experiência de consumo. Serviço completo (<i>full service</i>) corresponde a um serviço no qual o consumidor não tem envolvimento ou tem envolvimento limitado e assume um papel passivo. Já no <i>self-service</i> , o consumidor é responsável por todas as atividades envolvidas.	[26,27, 17]
Money Flow: caracteriza o processo de pagamento de uma transação iniciada em uma plataforma P2P SCC. Nesse sentido, muitas formas de pagamento podem ser categorizadas como uma transação C2C (<i>consumer-to-consumer</i>) usando métodos de pagamento típicos, como transferência bancária, dinheiro à vista, etc., ou C2B2C (<i>consumer-to-business-to-consumer</i>), onde a própria plataforma processa o pagamento e pode atuar como administradora/depositária.	[18]

Um das principais classes da ontologia é o conceito *d2s-owl:Platform*, que é usado para representar uma instância particular de uma plataforma P2P SCC. A Ontologia possui ainda algumas meta-informações diretas, como o ano de fundação (*d2s-owl:foundationYear*) ou o endereço web (*d2s-owl:hasUrl*) e também informações mais detalhadas, como *d2s-owl:MarketMediation* ou *d2s-owl:ConsumerInvolvement* (visto na tabela 3). A ontologia completa pode ser visualizada a seguir, na Figura 4.

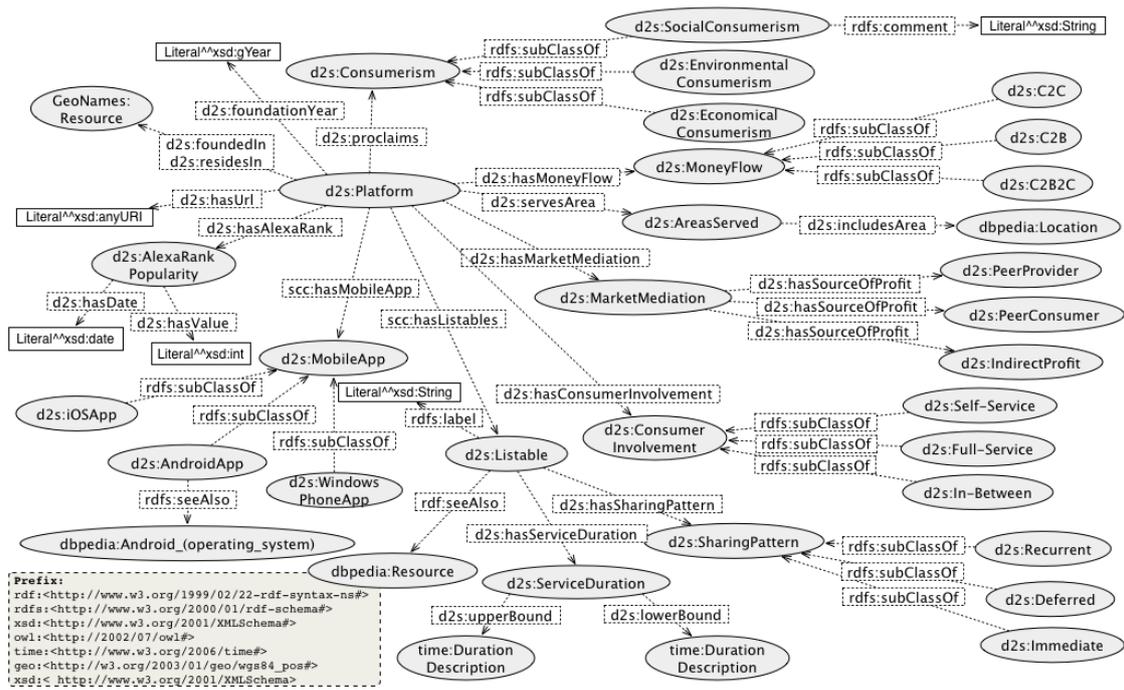


Figura 4: Ontologia do Discover2Share [4]

Outra contribuição desse trabalho foi o uso de alinhamento semântico com bases de conhecimento a fim de gerar valor agregado. Tal alinhamento faz parte do paradigma de Dados Conectados (*Linked Data*). Logo, referências a outros conjuntos de dados como DBpedia, YAGO e GeoNames foram consideradas, sempre que aplicáveis – para um conceito ou instância, dado que informações úteis estão disponíveis nesses conjuntos de dados específicos. Essa abordagem permitiu o enriquecimento semântico e o raciocínio básico, de modo que o conhecimento contido nesses conjuntos de dados externos pode ser usado para se fazer suposições e para se obter informações fundamentadas.

A tabela 4 apresenta uma visão geral dos conjuntos de dados, das informações contidas nos mesmos e também de sua utilização.

Tabela 4: Conjuntos de Dados Relevantes [adaptado de 4]

Conjunto	Informação	Utilização
Dbpedia ⁷ , YAGO ⁸	Informações Gerais e Taxonomias	Relacionamento, Sinônimos, Domínio
LinkedGeoData ⁹ , GeoNames ¹⁰	Informação Geográfica	Dados de Localização
Product Ontology ¹¹	Catálogo de Produtos	Descrições

O enriquecimento de informações e características de um determinado recurso, fornecido por uma plataforma; a obtenção de descrições e conceitos alternativos; a possibilidade de utilizar mapas hierárquicos e taxonomias para encontrar recursos relacionados, entre outros, são alguns dos benefícios obtidos com a utilização desses conjuntos de dados. Por exemplo, o link para a DBpedia ‘dbpedia.org/resource/Automobile’ pode ser utilizado para se inferir possíveis fabricantes de automóveis ou, pelo menos, obter uma descrição alternativa do conceito de automóvel.

Em linhas gerais, o uso da ontologia como linguagem de descrição, permitindo que os usuários possam tanto descrever as plataformas quanto suas ofertas de forma uniforme, juntamente com o alinhamento semântico com as bases de conhecimento, como o DBpedia, YAGO e outras, possibilitam uma descoberta melhorada das plataformas no Discover2Share.

⁷ DBpedia: <https://wiki.dbpedia.org>

⁸ YAGO: <http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/#c10444>

⁹ LinkedGeoData: <http://linkedgeodata.org>

¹⁰ GeoNames: <http://www.geonames.org/>

¹¹ Product Ontology: <http://www.productontology.org/>

4.4 Análise de Sentimentos ou Mineração de Opiniões

No passado, ao necessitar de uma opinião, os indivíduos perguntavam a seus amigos ou à sua família. No mundo dos negócios, quando precisavam de opiniões, as empresas conduziam pesquisas, grupos de estudo, entre outras iniciativas. Já os governos, quando desejam saber, por exemplo, o que se passava em outros países, monitoravam as mídias tradicionais, como jornais, radio e TV.

Atualmente, o advento da internet trouxe à tona o protagonismo dos indivíduos como produtores de conteúdo. Canais como blogs, fóruns e, sobretudo, as mídias sociais viabilizaram essa importante mudança, tanto na perspectiva dos próprios indivíduos, como na perspectiva das empresas e também dos Governos. Ao mesmo tempo em que esse fenômeno acontece, cresce a importância de tarefas como a Análise de Sentimentos, uma vez que todo esse conteúdo produzido representa uma valiosa fonte de dados e informações, tanto para os próprios indivíduos, quanto para as organizações.

A Análise de Sentimentos é o estudo, com a utilização de recursos computacionais, das opiniões, sentimentos, emoções e atitudes das pessoas extraídas de textos. Tal área, também conhecida como Mineração de Opiniões, tem grande importância tanto nos negócios quanto para a sociedade em geral. Como área de pesquisa, a Análise de Sentimentos é um dos tópicos com maior destaque dentro do campo de Processamento de Linguagem Natural, sendo estudada amplamente na mineração de dados, na mineração web e na recuperação de informação. Seu foco principal está nas opiniões que expressam, explícita ou implicitamente, sentimentos ou opiniões positivas e negativas. Tais opiniões normalmente dizem respeito a entidades que podem ser representadas por produtos, serviços, eventos, organizações ou indivíduos.

Em relação à terminologia, não há um consenso na literatura corrente sobre a diferença entre Análise de Sentimentos e a Mineração de Opiniões. Nesse sentido, há ainda outros termos relacionados, como:

- Análise de opinião;
- Extração de opinião;
- Mineração de sentimento;
- Análise subjetiva;
- Análise de afetividade;
- Análise de emoção; e
- Mineração de *reviews*.

As opiniões são um valioso ativo tanto para as empresas e organizações como também para os governos. Nesse sentido, esses são os primeiros interessados em saber como os seus consumidores ou público interagem/se relacionam com seus produtos e serviços, assim como na opinião sobre suas próprias marcas. Já os governos se interessam pelas opiniões da população a respeito de políticas públicas, leis e etc.

Segundo Atkinson et al [28], o crescimento no uso das mídias sociais gerou uma grande oportunidade para as pessoas publicarem as suas opiniões. Com isso, as organizações como um todo precisam ter em mãos um ferramental capaz de extrair a opinião de seus clientes de tais canais, interpretando-a e utilizando-a como importante insumo para o seu processo de tomada de decisão.

Para serem utilizadas como recurso para as organizações faz-se necessário que as opiniões estejam disponíveis de forma estruturada, uma vez que fontes de dados não estruturadas tendem a conter uma grande quantidade de informações e de conhecimentos implícitos [29].

Esse desafio, de classificar e analisar dados com o objetivo de identificar opiniões e sentimentos não é novo. Hatzivassiloglou e McKeown, em 1997, já tratavam da análise e extração de opiniões. Em 2002, Wiebe [30] lidava com a análise objetivando a identificação da subjetividade inclusa nas opiniões. Tais desafios evoluíram à medida que a internet também evoluiu, uma vez que as fontes desses dados também se modificaram ao longo dos anos. Num primeiro momento, textos publicados em sites eram o alvo da análise de sentimentos. Posteriormente, à medida que os jornais que antes eram impressos migraram para o meio digital, os mesmos se tornaram fonte de análise. Mais recentemente, as mídias sociais, onde os usuários passam a ter importante papel, conforme já mencionado, se tornaram um importante meio para a coleta de dados.

Em linhas gerais, a mineração de opiniões/análise de sentimentos tem como objetivo identificar, classificar e analisar opiniões e sentimentos através da polarização de uma sentença, documento ou texto como positiva ou negativa.

4.4.1 Níveis de Análise

A Análise de Sentimentos pode ser aplicada em três níveis de granularidade: nível de documento, nível de sentença ou nível de aspecto.

- **Nível de documento:** a classificação de sentimento em nível de documento consiste na tarefa de analisar o sentimento geral expresso em um documento inteiro [31; 32]. Normalmente, nesse nível, espera-se obter a opinião expressa em um documento a respeito de uma única entidade (empresa, produto, serviço). Logo, a presença de múltiplas entidades no documento torna necessária a análise em outro nível.
- **Nível de sentença:** nesse nível a análise visa determinar a opinião sobre uma entidade expressa em uma sentença – frase ou cláusula. Nesse nível a opinião

poderá ser positiva, negativa ou neutral (que pode ser considerada como ausência de opinião).

- **Nível de aspecto:** cada opinião é composta por um sentimento (positivo, negativo ou neutro) e um alvo (da opinião). Neste caso, um documento, e até mesmo uma sentença, pode conter opiniões sobre vários aspectos de várias entidades. Logo, temos nesse nível o máximo de detalhes sobre o sentimento exposto por um usuário.

4.4.2 Abordagens de Análise de Sentimento

Existem várias abordagens utilizadas na Análise de Sentimento. Nessa Seção algumas dessas abordagens serão apresentadas.

4.4.2.1 Abordagem baseada em Dicionário

Tal abordagem, também chamada de léxica ou linguística, faz uso de léxicos (dicionários) de sentimentos, previamente compilados com palavras ou expressões de sentimento associadas às suas polaridades.

Dentro dessa abordagem, um dos métodos mais utilizados é o de co-ocorrência entre alvo e sentimento, onde não é levado em consideração a ordem que os termos aparecem dentro do documento (*bag-of-words*) e nem as suas relações léxico-sintáticas. Nesse caso, basta que exista uma palavra expressando sentimento, cuja polaridade está contida em um léxico de sentimentos.

Esse método apresenta bons resultados quando a análise de sentimentos é realizada a nível de sentença ou em documentos curtos, uma vez que a palavra detentora do sentimento está normalmente próxima à entidade que qualifica.

4.4.2.2 Abordagem baseada em Aprendizado de Máquina

Nessa abordagem, as técnicas utilizadas tem como objetivo descobrir automaticamente regras gerais, existentes nos conjuntos de dados a serem analisados, que possibilitem a

extração de informações implicitamente representadas. Tais técnicas são divididas em dois tipos: as de aprendizado supervisionado e as de aprendizado não supervisionado.

Pegando como exemplo as técnicas supervisionadas, temos dois métodos predominantes: os de classificação e de regressão. Nesse sentido, o problema de classificação é separado em dois passos: aprendizagem de um modelo de classificação a partir de um corpus de treinamento previamente rotulado com as classes (negativo e positivo) e predição de novos textos com base no modelo resultado. Alguns dos algoritmos mais utilizados para essas tarefas são: Naïves Bayes, SVM (Support Vector Machine), Maximum Entropy e outros baseados em redes neurais.

A respeito dos métodos de regressão, usualmente a qualidade do modelo resultante da etapa de aprendizagem é medida levando-se em consideração a capacidade do modelo de prever corretamente (acurácia), a quantidade de instâncias previstas corretamente (precisão) e a quantidade de instâncias de uma classe analisada na classe correta (revocação).

As técnicas supervisionadas possuem algumas limitações, sendo a mais importante delas a necessidade de dados rotulados para treino. Dessa forma, o seu desempenho é afetado tanto pela qualidade quanto pela quantidade dos dados de treino disponíveis. Além disso, os conjuntos de treino são intimamente vinculados ao seu domínio.

4.4.2.3 Abordagem Estatística

Essa abordagem também é conhecida como não supervisionada e se baseia na premissa de que as palavras que denotam opiniões são frequentemente encontradas próximas umas às outras nos textos. Logo, se uma palavra ocorre com maior frequência junto a palavras positivas, então é provável que seja positiva. O mesmo critério vale para a ocorrência junto a palavras negativas, onde a mesma tende a ser negativa. Seguindo esse raciocínio, as palavras que ocorrem em igual frequência são classificadas como neutras. Algumas técnicas que fazem uso dessa abordagem são: Pointwise Mutual Information

(PMI), Semantic-orientation Latent Semantic Analysis (SO-LSA) e a Latent Dirichlet Allocation (LDA).

Em comparação às abordagens léxicas e por aprendizado de máquina, a abordagem estatística possui uma importante característica, que é ser menos suscetível à dependência do contexto, podendo, dessa forma, ser utilizada em conjunto com as demais a fim de complementá-las.

4.4.2.4 Abordagem Semântica

Essa abordagem guarda algumas semelhanças com a abordagem estatística. Entretanto, ela se distingue dessa segunda na forma como a polaridade é calculada: o cálculo é realizado em termos de alguma medida de distância entre os termos. Além disso, suas técnicas seguem o princípio de que palavras semanticamente próximas devem ter a mesma polaridade.

Nessa abordagem a polaridade também é obtida tendo como base palavras com sentimento positivo e negativo. A partir dessas é executado um processo de comparação ou expansão com a finalidade de determinar a polaridade dos termos. Esses processos normalmente mensuram a distância entre as palavras usando as relações como caminho ou através da contagem da frequência com que as palavras são associadas a sinônimos positivos ou negativos.

Embora também possa ser usada de forma a complementar outras abordagens, a abordagem semântica ainda é fortemente dependente de métodos manuais para a validação da polaridade atribuída, o que acaba dificultando a sua utilização.

4.4.3 Análise de Sentimentos Multi-Idioma

A maioria das abordagens e técnicas de análise de sentimentos, assim como os algoritmos existentes, foca em textos escritos no idioma inglês. Em paralelo, há alguns trabalhos que empregam abordagens adaptadas para idiomas específicos ou abordagens

multi-idioma. Essa realidade acaba limitando a quantidade de dados a serem utilizados, sobretudo quando o alvo da análise de sentimentos são os dados oriundos de conteúdo social, como nesse trabalho, por exemplo.

Dashtipour et al. [33] e Araújo et al. [34] tratam em seus trabalhos de abordagens multi-idiomas. Esse último estudo apresenta, inclusive, uma comparação de performance entre as técnicas específicas para o idioma inglês com uma abordagem multi-idioma que faz uso de textos traduzidos automaticamente para o idioma inglês.

Ao final, Araújo et al. [34] conclui que essa abordagem multi-idioma com dados traduzidos automaticamente, embora apresente um desempenho preditivo menor, pode ser competitiva se o método apropriado de análise de sentimento for devidamente escolhido. Na Seção 6.1.2.1 são apresentados os resultados da avaliação dessa abordagem realizada sobre os dados utilizados nesse trabalho, onde tal técnica foi empregada.

4.4.4 Apresentação dos Resultados da Mineração de Opiniões

Ao longo das últimas subseções foi apresentada a motivação para o uso da Análise de Sentimentos assim como a sua fundamentação. Nessa última subseção é tratada a apresentação dos resultados da análise de sentimentos/mineração de opiniões. O resultado da aplicação da análise de sentimentos, independente da técnica ou abordagem utilizada, consiste em se obter a classificação da polaridade de um texto, sentença ou aspecto. Nesse sentido, a exibição dos resultados da análise pode ser considerada como a última etapa dentro da análise de sentimentos. As abordagens de apresentação dos resultados da mineração de opiniões podem ser divididas em dois grupos: nível de documento e nível de característica.

Alguns exemplos de aplicação dessas abordagens são vistas a seguir:

4.4.4.1 Apresentação a nível de documento

Morinaga et al. [35], ao tratar da mineração da reputação de produtos na web, utilizou um *scatterplot* 2D (diagrama que faz uso de coordenadas cartesianas), ao qual deu o nome de mapa de posicionamento, para exibir o grupo de sentenças negativas ou positivas sobre um produto.

Gamon et al. [36] aborda em seu trabalho a mineração de opiniões a partir de textos livres. Como elemento de exibição do resultado da mineração é utilizada uma TreeMap (gráfico que usa retângulos aninhados para a exibição de dados hierárquicos) como interface para o usuário. Em tal gráfico são exibidos uma série de tópicos e o valor do sentimento médio estimado para cada tópico.

Chen et al. [37] trata da visualização de opiniões conflitantes. Nesse trabalho é utilizado um sistema de análise visual com múltiplas visões coordenadas onde são usadas árvores de decisão e gráficos de variação de termos.

Outro exemplo de sistema para visualização foi proposto por Draper [38]. Nesse sistema os usuários podem construir consultas (*queries*) visualmente e verem o resultado em tempo real.

4.4.4.2 Apresentação a nível de característica

Essa abordagem começou a ser utilizada a partir do desenvolvimento da mineração de opiniões baseada em aspectos. Alguns de seus exemplos são tratados a seguir.

Liu et al. [39] propôs um método para extrair opiniões a nível de características a partir de *reviews* de consumidores. Para exibir o resultado da análise foi utilizado um gráfico de barras aumentado. O objetivo da utilização desse elemento foi facilitar a comparação das opiniões por características.

Oelke et al. [40], ao tratar da análise visual de opiniões de *feedbacks* de consumidores, fez uso de relatórios de resumo, análise de cluster e mapa de correlação circular.

4.4.4.3 Abordagem híbrida

Em linhas gerais, a visualização ao nível de documento provê uma visão geral da opinião obtida, sem entretanto detalha-la. Já a visualização a nível de característica mostra a opinião específica, detalhada, sobre as características ou atributos de um produto ou serviço.

Mais recentemente, Wu [41] propôs uma abordagem híbrida para a visualização dos resultados da mineração de opinião. Em tal abordagem o elemento gráfico principal é o layout radial, composto pelo que foi chamado de “triângulo de opinião”. Esse radial tem como proposta integrar a visualização da opinião com múltiplas dimensões, incluindo informações demográficas, temporais e espaciais. Na Figura 5 são apresentadas as opiniões sobre o *feedback* de alguns usuários relacionado a alguns hotéis e que fazem parte de um experimento aplicado pelos autores.

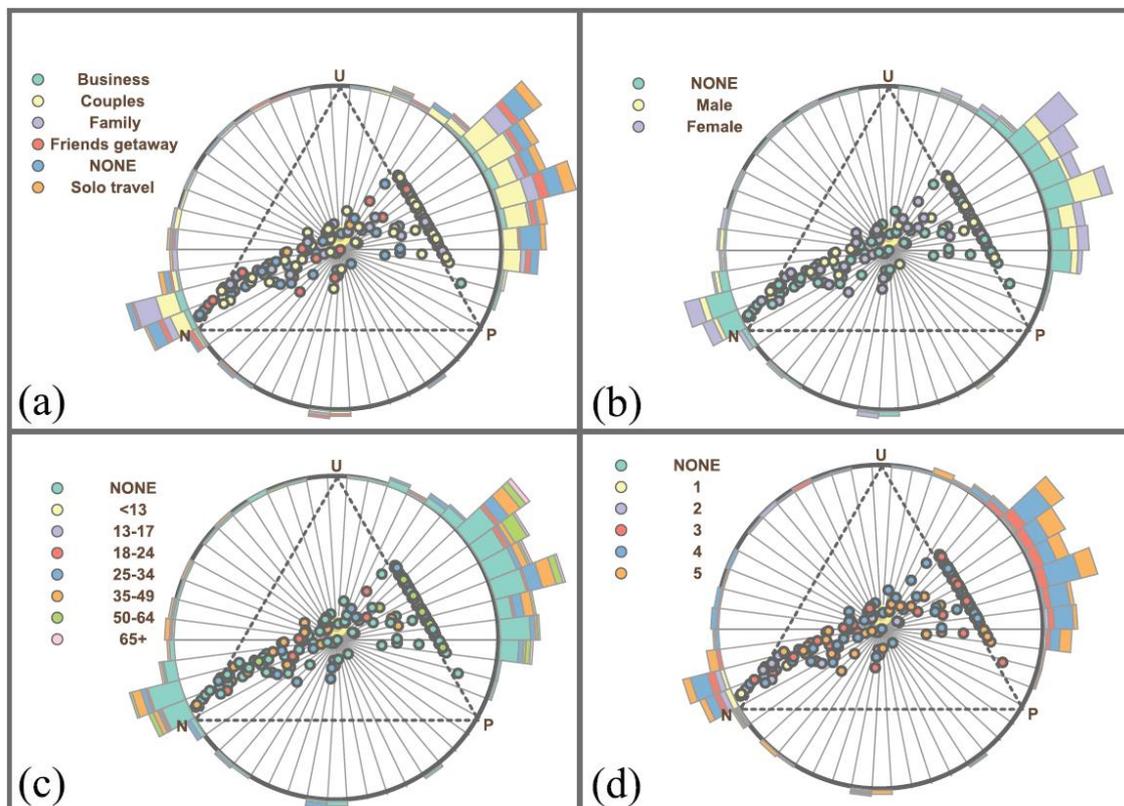


Figura 5: Radial contendo o Triângulo de Opiniões sobre Hotéis [41]

No elemento (a) são mostradas as polaridades (onde N significa negativas, U, incertas e P, positivas) agrupadas pelas motivações da viagem. Em (b) as polaridades são agrupadas por gênero. (c) apresenta o agrupamento por faixa etária e (d) pela classificação atribuída pelos usuários aos hotéis onde se hospedaram.

4.4.5 Ferramentas para Análise de Sentimentos - iFeel

Conforme Araújo et al. [42], o desenvolvimento da ferramenta (aplicação web) iFeel teve como objetivo principal a criação de uma ferramenta que permitisse a detecção de sentimentos em textos de qualquer formato, incluindo dados não estruturados oriundos de mídias sociais. Tal ferramenta permite o acesso a 13 métodos existentes de análise de sentimentos: Affin, Emolex, Emoticons, EmoticonsDS, NRC Hashtag Sentiment Lexicon, Opinion Lexicon, Sentiment140 Lexicon, Umigon, VADER, PANAS-t, Happiness Index, SenticNet e SentiStrength, descritos a seguir e utilizados neste trabalho.

A aplicação da análise de sentimentos sobre textos a partir da utilização dos 13 métodos mencionados concede ao iFeel um importante diferencial quando comparado a outras aplicações existentes, uma vez que é a única que permite o acesso a múltiplos métodos [42]. Além disso, esses métodos, que serão detalhados a seguir, fazem uso de um conjunto diferente de técnicas, baseadas em métodos léxicos ou em aprendizado automático.

4.4.5.1 Affin

Trata-se de uma lista de palavras em inglês classificada para valência (um léxico de sentimentos ou lista de palavras afetivas) com um número inteiro que varia entre menos 5 (negativo) a mais 5 (positivo). Tal lista foi rotulada manualmente por Finn Arup Nielsen entre 2009 e 2011. Embora existam outras listas de palavras com o mesmo intuito, como a ANEW, a motivação para a criação da Affin [43] foi desenvolver uma

lista de palavras especialmente para microblogs. Alguns exemplos de palavras e gírias não encontradas em outras listas mas muito comuns em textos postados na internet, e incluídas na Affin, são ‘WTF’ e ‘LOL’.

Os algoritmos que fazem uso de recursos léxicos utilizam uma base que contém termos previamente classificados, onde os mesmos são associados a notas positivas ou negativas. Logo, o processo de análise de sentimento utilizando um desses algoritmos consiste na comparação de cada termo presente no texto a ser analisado com os termos da base léxica. Ao final, o retorno individual de cada termo é somado. Se o resultado dessa soma for maior que zero, a polaridade do texto será positiva. Se menor que zero, negativa.

Pegando como exemplo a frase “The Airbnb platform is very stable and reliable.” e analisando com a Affin, teríamos o seguinte:

The{0} Airbnb{0} platform{0} is{0} very{0} stable{2} and{0} reliable{2}.{0}

Nessa frase, o valor entre chaves representa a pontuação de cada termo. Após o uso da Affin pode-se observar que não há palavras com pontuação negativa. Por outro lado, as palavras ‘stable’ e ‘reliable’ são positivas e possuem, cada uma, pontuação igual a 2. Logo, a soma da pontuação de toda a frase é igual a 4 e, portanto, tal frase tem polaridade positiva.

4.4.5.2 Emolex

O Emolex [44] é uma lista de palavras em inglês e suas associações com 8 emoções básicas (raiva, medo, antecipação, confiança, surpresa, tristeza, alegria e nojo) e dois sentimentos (negativo e positivo). Tal conjunto de palavras foi anotado manualmente através de *crowdsourcing*. Por também se tratar de um recurso léxico, sua aplicação é similar à da Affin, com o diferencial de apresentar, além da polaridade final do texto (positiva ou negativa) também a emoção identificada entre as 8 disponíveis.

4.4.5.3 Emoticons

Os *emoticons* são uma combinação de caracteres que, utilizados juntos, representam sentimentos ou emoções e tornaram-se bastante populares em programas de mensagem eletrônica e, sobretudo, nas redes sociais. Alguns exemplos de *emoticons* são:

- <3 que representa um coração;
- :) que representa um sorriso.

Na análise de sentimentos utilizando *emoticons* é realizada a extração de sentimentos em textos usando um grande conjunto de emoticons (onde também estão contidas as variações mais comuns de cada um) que expressam sentimentos positivos e negativos.

4.4.5.4 EmoticonDS

Tendo como motivação para sua criação o interesse em entender, a partir das mídias sociais, o impacto de fatores externos no sentimento agregado em uma sociedade, a EmoticonDS [45] foi desenvolvida a partir da análise e extração de sentimentos de 1,5 bilhão de tweets.

Inicialmente, foram observados os efeitos das condições meteorológicas e do tempo no sentimento agregado a fim de avaliar com que precisão o padrão individual se traduz em padrões aplicáveis a toda uma população, descobrindo-se ao final que o sentimento agregado segue distintos padrões climáticos, temporais e sazonais.

A exemplo da Emoticons, trata-se também de um recurso léxico, composto por uma lista de *tokens* (75,065 *tokens* únicos) e sua pontuação a nível de sentimento, com variação entre 0 e 1, representando tal escala a propensão de um *token* constar em *tweets* marcados com *emoticon* positivo ou negativo.

4.4.5.5 NRC Hashtag Sentiment Lexicon

O NRC Hashtag Sentiment Lexicon [46] trata-se também de um corpus léxico, criado a partir de posts do Twitter que contem *hashtags* expressando palavras de sentimento. Seu

diferencial para as demais listas de termos classificados de acordo com o sentimento que expressam, a partir de redes sociais, é estar direcionado às palavras precedidas pelo símbolo *hash* (#), que no Twitter recebem o nome de *hashtags*.

Tomando como exemplo a frase “*Feeling left out... #sadness*”, percebe-se a existência de uma palavra que expressa sentimento associada a um hash – a hashtag ‘#sadness’. Nesse exemplo, utilizando o NRCHashtag, a polaridade da frase em questão é negativa.

4.4.5.6 Opinion Lexicon

O Opinion Lexicon [47] também faz uso de um léxico para classificar a polaridade de textos – nesse caso, de sentenças. Para tal, três sub-tarefas são executadas: primeiro, um conjunto de adjetivos (que são palavras normalmente utilizadas para expressar opiniões) é identificado usando um método de processamento de linguagem natural. A seguir, para cada palavra contendo opinião, é determinada a sua orientação semântica, ou seja, se é negativa ou positiva. Para essa sub-tarefa é utilizada a WordNet¹². Por fim, o algoritmo determina a orientação da opinião de cada sentença.

4.4.5.7 Sentiment140 Lexicon

Desenvolvimento pelos mesmos autores do NRC Hashtag e do Emolex, o Sentiment140 Lexicon [48] foi gerado a partir de *tweets* com *hashtags* ‘palavra-sentimento’ e *tweets* com *emoticons* com a utilização dos classificadores criados para a competição internacional organizada pela Conferência em Exercícios de Avaliação Semântica (SemEval-2013). Trata-se, portanto, de mais um léxico baseado na associação de ‘palavras-sentimento’.

¹² WordNet: <https://wordnet.princeton.edu/>

4.4.5.8 Umigon

O Umigon [49] é uma aplicação web que provê um serviço de detecção de sentimentos em tweets (textos postados no Twitter). Entre suas funcionalidades destaca-se a indicação de características semânticas adicionais presentes nos tweets, como a indicação de tempo ou marcadores de subjetividade.

Integrante da família de classificadores de sentimentos baseados em léxicos, o Umigon é especificamente designado para detectar sentimento (positivo, negativo ou neutro) em *tweets*. Seu motor de detecção de sentimento consiste em quatro partes:

- Detecção de características semânticas no *tweet* inteiro, onde ‘*Smileys*’ e onomatopeias ganham especial atenção;
- Avaliação de *hashtags*;
- Decomposição do *tweet* em uma lista de ‘*n-grams*’ (chegando ao máximo de *4-grams*) e comparação de cada ‘*n-gram*’ com os termos do léxico, onde, havendo igualdade, a heurística é aplicada;
- Série final de heurísticas no nível de todo o *tweet*, aproveitando as características semânticas detectadas nas demais etapas. Ao final, um sentimento único é atribuído ao *tweet*.

Ao todo, o Umigon faz uso de quatro léxicos contendo termos com ‘tom positivo’, ‘tom negativo’, ‘força do sentimento’ e ‘negações’ para a análise de sentimentos. Tais léxicos foram criados manualmente e continuam sendo expandidos. Ainda em relação aos termos contidos nos léxicos, cada um deles é acompanhado por uma heurística e uma regra de decisão.

As tabelas 5 e 6 apresentam exemplos de dois termos, o primeiro expressando sentimento negativo e o segundo sentimento positivo, contido nos léxicos do Umigon.

Tabela 5: Exemplo de termo com sentimento negativo contido nos léxicos do Umigon
[Adaptado de 49]

Termo	Sadfaced
Heurística	None
Regra de Decisão	012

Tabela 6: Exemplo de termo com sentimento positivo contido nos léxicos do Umigon
[Adaptado de 49]

Termo	Satisfied
Heurística	!isImmediatelyPrecededByANegation
Regra de Decisão	011 012

No primeiro caso, se o *tweet* contem o termo ‘sadfaced’, ele receberá durante a análise pelo Umigon o código 012, que significa um sentimento negativo.

No segundo caso, se o termo ‘satisfied’ estiver presente no *tweet*, a heurística ‘!isImmediatelyPrecededByANegation’ é aplicada, através de um método que checa se o termo imediatamente anterior a ‘satisfied’ no *tweet* é uma negação ou não. A partir do retorno (falso ou verdadeiro) desse método é determinada a regra de decisão a ser aplicada. Tal regra é composta por uma escolha entre dois binários ‘011’ (caso ‘satisfied’ não for precedido por uma negação) ou ‘012’ (caso ‘satisfied’ for precedido por uma negação). O primeiro binário expressa um sentimento positivo e o segundo, um sentimento negativo.

4.4.5.9 VADER

VADER [50] - *Valence Aware Dictionary for sEntiment Reasoning* - é um modelo para análise de sentimentos que faz uso de uma lista de recursos léxicos (juntamente com as medidas de intensidade de sentimento) combinados com cinco regras que incorporam convenções gramaticais e sintáticas para expressar e enfatizar a intensidade do sentimento. O léxico de sentimentos gerado na criação do VADER está especificamente sintonizado ao contexto de sentimentos expressos em textos de *microblogs*.

A seguir são demonstrados dois exemplos de análise de sentimentos utilizando o VADER:

Sentença 1: “VADER is smart, handsome, and funny.”

Análise: {'neg': 0.0, 'neu': 0.254, 'pos': 0.746, 'compound': 0.8316}

Polaridade da Sentença 1: positiva.

Sentença 2: “Make sure you :) or :D today!”

Análise: {'neg': 0.0, 'neu': 0.294, 'pos': 0.706, 'compound': 0.8633}

Polaridade da Sentença 2: positiva.

Como pôde ser visto, a sentença é avaliada e recebe uma pontuação para as polaridades negativa, neutra e positiva. A composição das três (*'compound'*) é a pontuação final da sentença. Se essa pontuação for maior que 0.5, então a sentença é positiva. Se a pontuação variar entre -0.5 e 0.5, então é neutra. Abaixo de -0.5, é negativa.

4.4.5.10 PANAS-t

A PANAS-t [51] é uma escala psicométrica de onze sentimentos, baseada na versão expandida da conhecida escala PANAS-x (*Positive Affect Negative Affect Scale*) e adaptada para medir sentimentos nos textos curtos da rede social Twitter.

A PANAS-x, que é uma versão expandida da escala original PANAS adaptada para o contexto da Web mede onze afetos específicos: medo, tristeza, culpa, hostilidade, timidez, fadiga, surpresa, jovialidade, auto-confiança, serenidade e atenção. A tabela 7 resume os itens que compõem a escala PANAS-x.

Tabela 7: Termos contidos na escala PANAS-x [Adaptado de 51]

Estado afetivo	Palavras relacionadas (em inglês)
Medo	afraid, scared, frightened, nervous, jittery, shaky.
Hostilidade	angry, hostile, irritable, scornful, disgusted, loathing.
Culpa	guilty, ashamed, blameworthy, angry at self, disgusted with self, dissatisfied with self.
Tristeza	sad, blue, downhearted, alone, lonely.
Jovialidade	happy, joyful, delighted, cheerful, excited, enthusiastic, lively, energetic.
Auto-confiança	proud, strong, confident, bold, daring, fearless.
Atenção	alert, attentive, concentrating, determined.
Timidez	shy, bashful, sheepish, timid.
Fadiga	sleepy, tired, sluggish, drowsy.
Serenidade	calm, relaxed, at ease.
Surpresa	amazed, surprised, astonished.

Embora adaptado do PANAS-x, o PANAS-t possui algumas diferenças em relação ao mesmo: o PANAS-x fornece uma série de adjetivos que um entrevistado deve

quantificar, enquanto o PANAS-t colhe sentimentos espontâneos expressos no Twitter. Além disso, o PANAS-x precisa que o entrevistado exponha o que e o quanto ele está sentindo, enquanto o PANAS-t filtra sentimentos individuais sobre eventos específicos. Essas diferenças são importantes pois poderiam, por exemplo, afetar como os sentimentos são medidos em ambas escalas.

A análise de sentimentos utilizando o PANAS-t consiste em avaliar um conjunto de textos (tweets) e medir a variação de cada estado afetivo, que pode ir de -1 a 1. Ou seja, um determinado texto onde o estado afetivo ‘medo’ foi medido como 0 indica que não houve nem aumento nem diminuição do sentimento em questão. Em outras palavras, tal sentimento não foi identificado no texto analisado.

4.4.5.11 Happiness Index

O Happiness Index [52] é uma escala de sentimento que faz uso das normas afetivas das palavras em inglês (ANEW - Affective Norms of English Words). ANEW é uma coleção de aproximadamente 1000 palavras comumente usadas associadas com suas dimensões afetivas de valência, excitação e dominância.

Ao final da análise de sentimentos utilizando o método Happiness Index cada texto obtém uma pontuação que varia de 1 a 9. A polaridade do texto classificado entre 1 e 4 é considerada como negativa e a de 6 a 9 como positiva.

4.4.5.12 SenticNet

O SenticNet [53] é um método que explora técnicas de inteligência artificial e de semântica Web. Seu objetivo é inferir a polaridade dos conceitos de senso comum de textos em linguagem natural em um nível semântico, em vez de sintático. Tal método utiliza ainda técnicas de Processamento de Linguagem Natural (NLP – *Natural Language Processing*) para criar a polaridade de aproximadamente 14.000 conceitos.

4.4.5.13 SentiStrength

O SentiStrength [54] é um programa que foi desenvolvido para a detecção do peso de

sentimentos expressos em textos oriundos de conteúdo social. O SentiStrength faz uso de uma abordagem léxica que explora uma lista de termos baseados em sentimentos e possui regras para ligar com métodos de linguística padrão e conteúdo social para expressar sentimentos como emoções, pontuação exagerada e erros ortográficos deliberados.

A aplicação do SentiStrength pode ser vista nesse exemplo:

Texto: “Airbnb is very coooolllll”

Resultado: positive strength 4 and negative strength -1

Base lógica para a classificação: airbnb is very coooolllll[2] [+0.6 spelling emphasis] [+1 booster word] [sentence: 4,-1]

Polaridade: positiva.

A escala de sentimentos positivos na SentiStrength varia entre 1 (não positivo) a 5 (extremamente positivo) e a escala de sentimentos negativos vão de -1 (não negativo) a -5 (extremamente negativo).

4.5 Processos Intensivos em Conhecimento

O processo de escolha de plataformas no Discover2Share, conforme já mencionado, é caracterizado como intensivo em conhecimento. Esse tipo de processo é dependente do conhecimento embutido nas pessoas que o executam e nas atividades e tarefas que os compõem [55].

Maldonado [56] cita as principais características dos Processos Intensivos em Conhecimento (PIC):

- A sequência de atividades e a agregação de valor são dependentes da aquisição e uso do conhecimento;
- Tais processos evoluem à medida em que são executados, se baseando no conhecimento adquirido;

- Suas especificações são pouco estruturadas e o fluxo de atividades é dinâmico;
- Os atores possuem forte influência no resultado do processo e têm a possibilidade de reutilizar e adaptar conhecimento de outros domínios e de diferentes níveis de expertise;
- Há alta imprevisibilidade na execução das atividades;
- Por ser atualizado constantemente, o tempo de vida útil do conhecimento envolvido no processo é geralmente curto e normalmente requer bastante tempo para ser adquirido;
- Por não seguirem regras estruturadas de trabalho, as suas medidas de desempenho são difíceis de serem estabelecidas e mensuradas;
- O suporte das tecnologias de informação a esses processos costuma não ser sofisticado, já que os mesmos dependem intensivamente da socialização e da troca informal de conhecimento;
- O seu custo de gestão costuma ser elevado.

Um exemplo de processo intensivo em conhecimento pode ser visto na Figura 6, onde é apresentado o modelo do processo atual de escolha de plataforma no diretório Discover2Share, realizada neste trabalho. Em tal processo, desde o seu primeiro passo, ou seja, a escolha por avançar selecionando uma região geográfica ou um tipo de recurso, já implica conhecimentos prévios do usuário, pois embora a navegação por região geográfica possa ser intuitiva, a navegação por tipo de recurso requer saber antecipadamente o que são esses recursos ou ainda um conhecimento aprofundado nas plataformas P2P SCC. O mesmo raciocínio se aplica nos passos de escolha e navegação

pelos atributos, já que alguns deles são bastante específicos e também exigem algum conhecimento prévio.

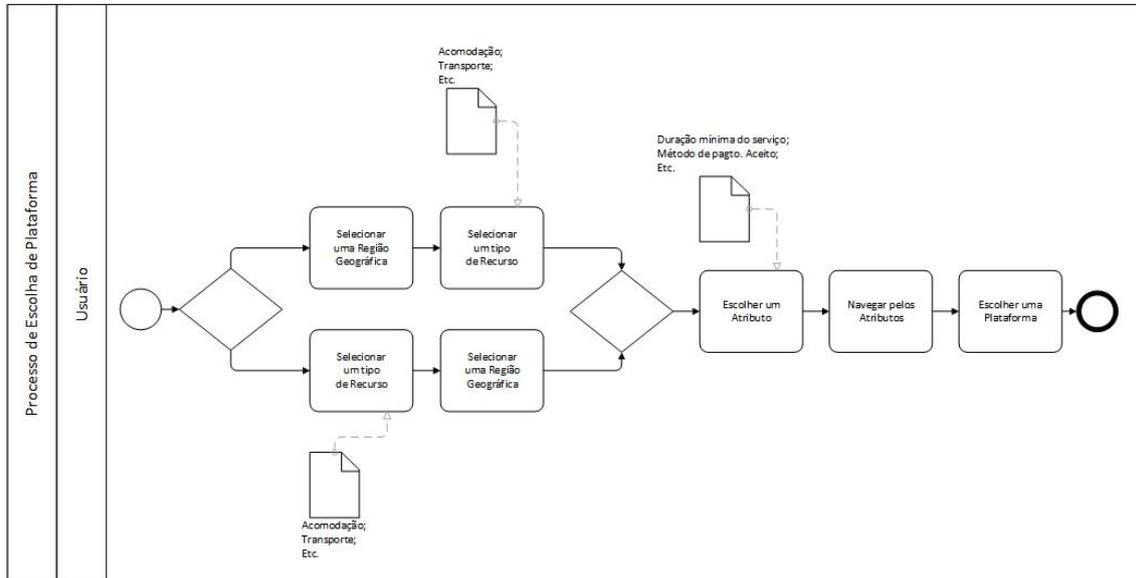


Figura 6: Processo Intensivo em Conhecimento – Escolha de plataforma no Discover2Share [Elaboração Própria]

5. Método da Proposta de Solução

Este trabalho propõe um método, que pode ser visto na Figura 7, baseado em técnicas utilizadas em processos de descoberta de conhecimento – mais precisamente, na análise de sentimentos/mineração de opiniões, para a coleta, pré-processamento, análise e apresentação dos dados para enriquecimento do Discover2Share. Tal método, que será detalhado a seguir, é composto por quatro fases: Coleta dos Dados, Pré-processamento, Análise dos Dados e Apresentação.

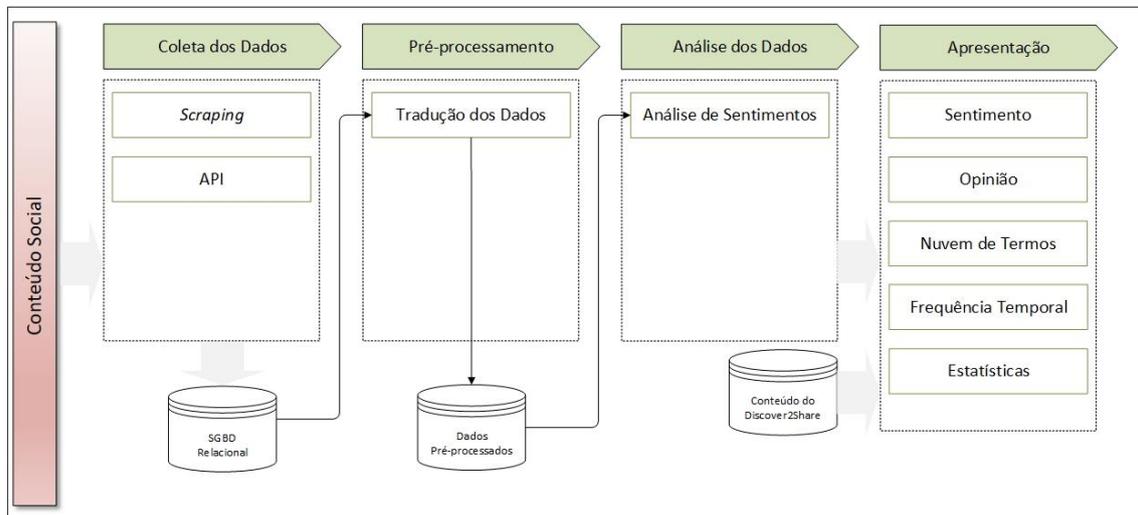


Figura 7: Visão geral do método para enriquecimento do Discover2Share

[Elaboração própria]

5.1 Fase 1: Coleta dos Dados

Na primeira fase é realizada a coleta do conteúdo social, sendo obtidos os dados a serem utilizados a partir de posts do Twitter, do Facebook e de avaliações publicadas nas próprias plataformas. O critério empregado para filtrar os posts de interesse nas redes sociais foi utilizar os próprios nomes das plataformas como palavras-chave nas buscas. Sobre a aquisição em si, há algumas formas de coletar os dados em questão. No caso das plataformas, foi desenvolvido um programa (*parser*) que coleta os dados das APIs

fornecidas pelas mesmas. No caso do Twitter, também existe uma API oficial que permite o acesso ao conteúdo público disponível. Entretanto, essa API tem uma limitação importante, pois só permite o acesso a dados recentes – publicados nos últimos sete dias. A mesma situação é encontrada no Facebook. Logo, foi preciso contornar tais limitações, já que se pretendia trabalhar com uma quantidade maior de informações. Para isso foram desenvolvidos scripts de ‘web crawling / scraping web’, cuja principal função é coletar os dados das redes sociais em questão sem fazer uso de suas APIs, ou seja, acessando e coletando os dados diretamente de suas páginas.

Todas as atividades de coleta de dados foram automatizadas com o uso de programas escritos em Java e também em Python. Além disso, a coleta foi agendada para ser executada de forma autônoma e precisou de alguns dias para ser concluída, já que as limitações anteriormente citadas não permitem a coleta em massa dos dados. Cabe citar ainda que essas limitações impedem, por exemplo, que essa abordagem seja utilizada em tempo real a partir das pesquisas dos usuários. Por outro lado, os scripts usados podem ser configurados para execução contínua, aumentando significativamente a quantidade de dados disponíveis.

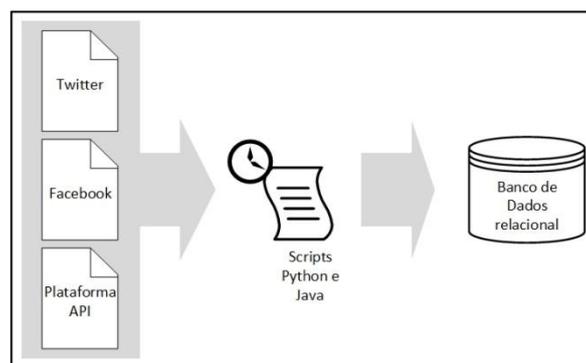


Figura 8 – Coleta de Dados [Elaboração própria]

Ao final dessa fase, os dados coletados são armazenados em um banco de dados relacional. No escopo dessa proposta não está prevista a integração física do banco de dados aqui utilizado com o existente no Discover2Share. Logo, tratam-se de bases de dados independentes.

5.2 Fase 2: Pré-processamento

Normalmente, nos processos de mineração de dados, mineração de opiniões e outros de descoberta de conhecimento a partir de textos, é comum haver uma fase de pré-processamento. Tal fase é composta por atividades que têm como objetivo melhorar a qualidade dos textos adquiridos, assim como organizá-los para posterior análise, através de um conjunto de transformações aplicados sobre os mesmos. Alguns exemplos das atividades tipicamente executadas no pré-processamento de dados são: *tokenização* (que é a divisão do texto em palavras, frases ou outros elementos chamados *tokens*); *stemming* (que é a normalização morfológica das palavras contidas nos textos); remoção de *stopwords* (tratam-se de palavras irrelevantes, normalmente, mas não somente, representadas pelas preposições, artigos e conjunções); e *POS-Tagging* (que é a identificação das classes gramaticais).

Nesse trabalho optou-se por não fazer uso dessas técnicas. Tal decisão se apoiou tanto nas fontes e nos dados colhidos - que são em sua maioria textos curtos, formados por poucas sentenças e onde normalmente são encontrados caracteres especiais, abreviaturas e *emoticons*, quanto nos algoritmos usados na análise de sentimentos, e que serão discutidos posteriormente. Primeiro, porque a eventual exclusão de alguns dos elementos citados pode acabar dificultando a identificação dos sentimentos expressos nos textos. Segundo, porque os algoritmos estão preparados para lidar com este tipo de dados e com os elementos neles encontrados.

Com isso, o único processo executado nessa fase foi o de tradução dos textos, que ainda não estavam nesse idioma, para o inglês com o uso da ferramenta *Google Translation API*¹³.

Para realizar a tarefa de tradução foi desenvolvido um programa que recupera os dados a partir do banco de dados, os submete à API e a seguir grava os textos traduzidos na base de dados. Cabe ressaltar que os textos originais, em seus idiomas originais, foram salvos juntamente com suas versões traduzidas.

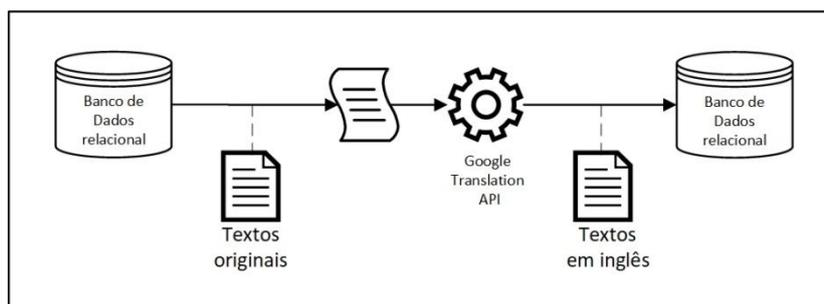


Figura 9 – Pré-processamento dos Dados [Elaboração própria]

5.3 Fase 3: Análise dos Dados

Nessa fase são aplicados os processos de análise de sentimentos sobre os dados coletados e pré-processados nas etapas anteriores.

Como mostrado na Figura 10, a ferramenta iFeel [42] foi utilizada para a análise dos textos. Além de disponível online, é possível utilizar essa ferramenta através de seu código-fonte, disponibilizado pelos autores na linguagem de programação Java. Essa segunda opção foi utilizada.

¹³ Google Translation API: <https://cloud.google.com/translate/>

O processo de mineração de opinião consiste na submissão individual de cada texto à ferramenta iFeel, que implementa 13 algoritmos de análise de sentimentos (Affin, Emolex, Emoticons, EmoticonsDS, NRC Hashtag Sentiment Lexicon, Umigon, VADER, PANAS-t, Happiness Index, SenticNet e SentiStrength), detalhados na Seção 4.4.5, identifica os sentimentos contidos nos textos, i.e., sua polaridade, que pode ser positiva, neutra ou negativa.

Os 13 algoritmos em questão (que não podem ser escolhidos pelo usuário, ou seja, todos são executados) podem gerar diferentes classificações de polaridade para cada texto. Dessa forma, é considerado como sentimento expresso no texto a polaridade preponderante (a maioria) entre as 13 obtidas.

Todas as atividades acima foram realizadas com a utilização de programas em Java, desenvolvimentos especificamente para essas tarefas, que foram integrados à ferramenta iFeel para realizar a submissão e análise dos textos e também o armazenamento do seu resultado ao final, onde a polaridade obtida de cada texto é armazenada no banco de dados juntamente com a referência para o respectivo texto. Futuramente, a inclusão de novos algoritmos, seja pela atualização da ferramenta iFeel ou pelo desenvolvimento de algoritmos próprios, poderá ser feita através da adaptação desses scripts desenvolvidos.

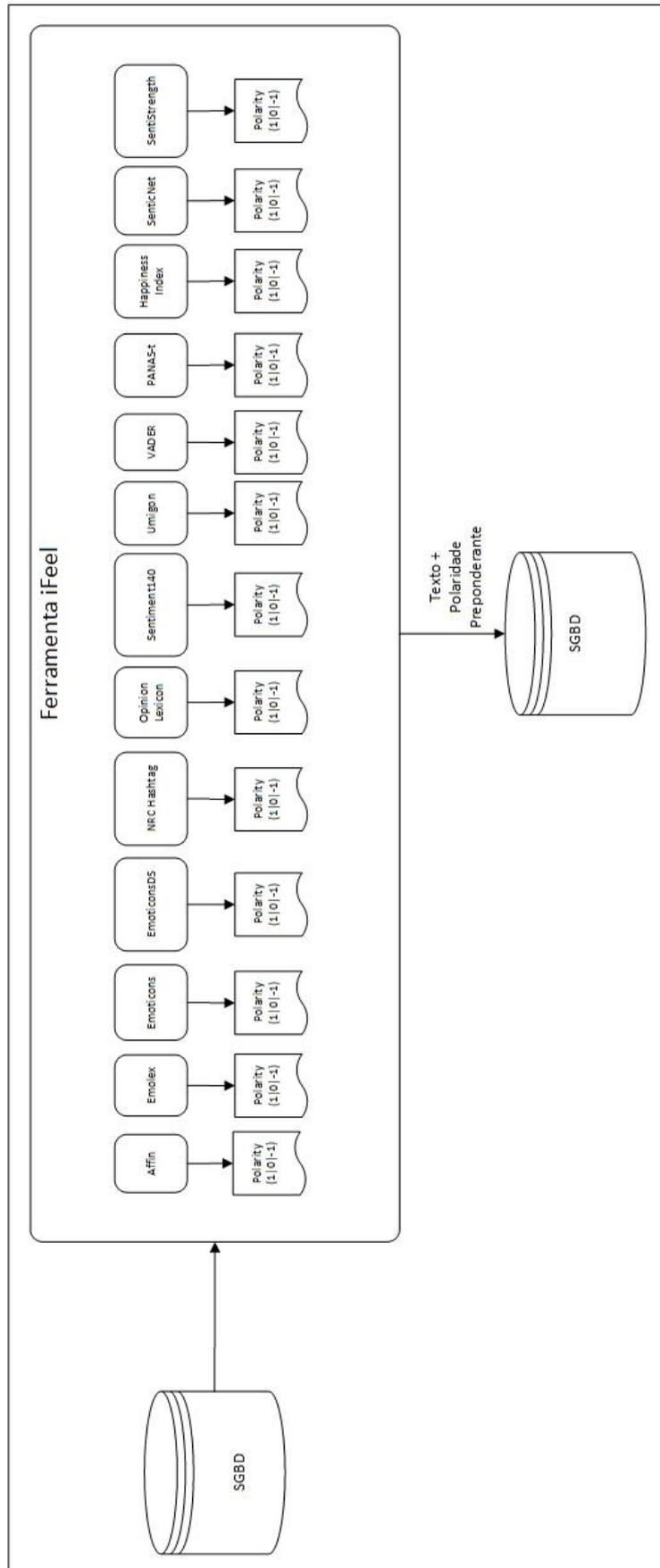


Figura 10: Análise de Sentimentos utilizando a ferramenta iFeel [Elaboração própria]

5.4 Fase 4: Apresentação

Nas fases anteriores os dados foram coletados, armazenados, pré-processados e também analisados. Ao final dessas três fases, encontram-se armazenados na base de dados: os dados originais; os textos traduzidos para o inglês; e os sentimentos expressos em cada texto – sua polaridade. Logo, a última fase dessa metodologia consiste em apresentar todas essas informações para o usuário.

Para a apresentação são usadas séries temporais, tabelas, nuvens de tags e grids. Além dessas informações, que são resultado da aplicação da metodologia ora proposta, também são apresentadas as informações existentes atualmente no Discover2Share, através de uma tabela. Todos esses elementos gráficos foram incluídos em uma página Web, desenvolvida para tal finalidade com o uso de HTML e da biblioteca *javascript* D3.js¹⁴.

A partir da apresentação é fornecida uma visão-geral a respeito das plataformas, sendo fornecidas informações úteis em relação à sumarização das polaridades identificadas, à quantidade de dados obtida para cada plataforma, os termos mais mencionados em cada uma, assim como os textos originais e suas polaridades.

¹⁴ D3 Data-Driven Documents: <https://d3js.org/>

6. Aplicação dos Estudos de Caso

Nesse Capítulo é apresentada a aplicação dos dois Estudos de Caso realizados para avaliar a proposta de solução. O primeiro, um experimento onde foram coletados dados de três plataformas P2P SCC: 9Flats, Airbnb e HomeAway. O segundo, a realização de um questionário on-line.

6.1 Aplicação em um caso real do Método da Proposta de Solução

Foi realizado um experimento onde o método da proposta de solução, apresentado no Capítulo 5, foi aplicado em um caso real, onde foram coletados dados de três plataformas P2P SCC: 9Flats, Airbnb e HomeAway. Essas plataformas fornecem recursos do tipo ‘acomodação’ e foram escolhidas por possuírem uma quantidade considerável de posts no Twitter e no Facebook. Além disso, as duas primeiras possuem uma API (Interface de Programação de Aplicativos, do inglês *Application Programming Interface*) que permite o acesso aos comentários postados por seus usuários. Embora apresentem uma visão particular sobre as plataformas, onde os dados provenientes das redes sociais tratam mais das próprias plataformas e os das APIs sejam mais direcionados às acomodações alugadas através das plataformas, os dados foram analisados em conjunto, gerando assim pequena influência no resultado final.

A estrutura dessa Seção está dividida da seguinte forma: inicialmente são descritos os dados utilizados no experimento. Na Seção seguinte, a etapa de processamento é tratada. A seguir, a abordagem é aplicada e seus resultados são demonstrados.

6.1.1 Aquisição dos Dados

Foram selecionadas três fontes de dados para a aplicação da abordagem ora proposta: Twitter, Facebook e as próprias plataformas. Em relação ao Twitter, foram utilizados os próprios nomes das plataformas como termos de pesquisa, ou seja: ‘9Flats’, ‘Airbnb’ e ‘HomeAway’. A partir disso, os scripts utilizados foram instanciados para coletar todos os *posts* que continham tais termos.

No Facebook a coleta de dados foi realizada a partir dos posts públicos disponíveis nas páginas oficiais das três plataformas. Logo, aqui temos uma restrição para a utilização do Facebook como fonte de dados, já que é necessário que a plataforma a ser analisada possua uma página em tal rede social.

Na terceira fonte de dados usada, as próprias plataformas com as avaliações publicadas pelos seus usuários, foi possível coletar dados das plataformas 9Flats e Airbnb, que disponibilizam uma API que permite a aquisição de tais dados. Já a HomeAway não possui tal recurso.

Todos os scripts utilizados para a coleta de dados foram instanciados para coletarem dados, nas três fontes usadas, relativos aos meses de janeiro e fevereiro de 2017. A Tabela 8 mostra um descritivo sumarizado dos dados coletados.

Tabela 8. Visão Geral dos Dados Coletados em dois meses

Fonte de Dados	Airbnb			9flats			HomeAway		Total
	API	Twitter	Facebook	API	Twitter	Facebook	Twitter	Facebook	
Volume	2,575	131.478	2,913	39	59	0	4,054	217	141.335
Proporção	1.82%	93.03%	2.06%	0.03%	0.04%	0%	2.87%	0.15%	100%

6.1.2 Pré-processamento

Conforme descrito na Seção 5.2, após a coleta dos dados os mesmos foram submetidos a um processo de tradução para o inglês, quando já não nesse idioma. Com a utilização da ferramenta Google Translation API, os dados foram recuperados do banco de dados,

traduzidos e armazenados novamente, sendo mantidos tanto os textos originais quanto os novos. Todo esse processo foi automatizado e, ao seu final, 39617 textos foram traduzidos para o inglês.

6.1.2.1 Avaliação do Método Multi-Idioma

A abordagem multi-idioma (análise de sentimentos aplicada sobre textos automaticamente traduzidos para o idioma inglês), melhor discutida na Seção 4.4.3, é um dos pilares da abordagem utilizada. Logo, uma das preocupações adicionais nesse trabalho foi verificar a validade de tal método. Para isso foram realizados alguns experimentos com o intuito de medir a precisão do método. Tais experimentos são descritos a seguir.

Inicialmente foi definida a medida a ser utilizada para aferir a validade do método: precisão. Nesse sentido, é considerada como precisão no escopo dessa avaliação o percentual de acertos obtidos pelo método, em comparação com a classificação realizada manualmente, dado pela fórmula:

$$P = (T^a / T^t) * 100, \text{ onde}$$

P = Precisão, em termos percentuais;

T^a = Total de acertos (descrito a seguir); e

T^t = Total de Textos.

A seguir, foi calculado o total de acertos do método. Para tal, foram utilizados 80 textos dentre os coletados e que estavam no idioma português. Desses textos:

- 60 foram provenientes da plataforma Twitter, divididos em grupos de 20 registros por cada polaridade (positivo, negativo e neutro). Tal polaridade foi atribuída pelo método sobre os textos automaticamente traduzidos para o inglês;

- 20 foram provenientes da plataforma Facebook, sendo todos de polaridade positiva.

Cabe ressaltar que não existiam mais textos no idioma português oriundos dos dados das próprias plataformas ou contendo polaridades negativas ou neutras vindos do Facebook.

A seguir, todos os dados acima – os textos e as respectivas polaridades, agrupados por origem dos dados, foram tabulados e então cada um deles teve sua polaridade aferida manualmente.

Por fim, foram obtidos os resultados da medição da precisão a partir da comparação entre a polaridade atribuída pelo método sobre os textos em inglês versus a polaridade atribuída manualmente sobre os textos originais, em português. As tabelas 9 e 10 mostram os resultados, que serão discutidos a seguir.

Tabela 9. Percentual de Acertos em relação aos dados totais

Precisão	Twitter			Facebook		
	Positivas	Negativas	Neutras	Positivas	Negativas	Neutras
Positivas	30%	10%	5%	100%	-	-
Negativas	5%	50%	10%	-	-	-
Neutras:	65%	40%	85%	-	-	-
Média	55%			100%		

Tabela 10. Percentual de Acertos em relação aos dados com polaridade definida

Precisão	Twitter	
	Positivas	Negativas
Positivas	86%	17%
Negativas	14%	83%
Média	85%	

A Tabela 9 mostra o percentual de acertos por polaridade, comparando a classificação manual com a automática. Os cruzamentos Positivas x Positivas, Negativas x Negativas

e Neutras x Neutras mostra a precisão obtida. Ou seja, em relação ao Twitter, a precisão de classificação dos dados positivos foi de 30%. Nessa fonte de dados, a classificação manual indicou que 5% dos textos automaticamente classificados como positivos possuíam polaridade negativa e 65% neutra. A classificação dos dados negativos apontou uma precisão de 50%. Aqui, a classificação manual apontou que 10% dos textos automaticamente classificados como negativos eram positivos e 40% neutros. Em relação aos textos automaticamente classificados como neutros, a precisão obtida foi de 85%. Sendo apontados pela classificação manual como 5% positivos e 10% negativos os textos automaticamente apontados como neutros. Ao final, a precisão medida (média dos percentuais de acerto em cada polaridade, ou seja, 30% das positivas + 50% das negativas + 85% nas neutras) da fonte de dados Twitter ficou em 55%.

Em relação ao Facebook, onde havia apenas registros automaticamente classificados como positivos, a precisão foi de 100%.

Finalmente, a Tabela 10 apresenta a precisão quando considerados apenas os dados com polaridade identificada, ou seja, sendo desconsiderados os dados neutros, que também indicam a ausência de sentimento. Nesse caso, a média final da precisão foi de 85%.

Os valores finais de Precisão obtidos indicam que o método é viável e fornece resultados aceitáveis. Entretanto, demonstram ser necessários estudos mais aprofundados, em trabalhos futuros, tanto na aferição de outras medidas, quanto na aferição sobre os textos originalmente em inglês e até mesmo da aferição dos algoritmos usados via plataforma iFeel.

6.1.3 Apresentação dos Resultados

A abordagem proposta nesse trabalho foi utilizada para coletar opiniões, através de posts e *reviews* de diferentes fontes, relativas a três plataformas P2P SCC. Após a coleta

em si, os dados foram pré-processados e analisados, onde a polaridade de cada texto foi obtida e armazenada no banco de dados. Dessa forma, a última etapa dessa abordagem consiste em apresentar ao usuário final todos os dados coletados e, sobretudo, o resultado das análises realizadas. Para facilitar a visualização foi desenvolvido um *dashboard* que sumariza todas as informações usando diferentes elementos gráficos, como frequências temporais, visto nas Figuras 14,15 e 16, barras, visto na Figura 11, nuvens de tags, Figura 17, e grid, Figura 18. Cada um desses elementos é descrito a seguir.

6.1.3.1 Visão Geral dos Dados Disponíveis

Inicialmente é mostrada ao usuário uma visão geral dos dados disponíveis. Para isso foi utilizado um gráfico de barras, que facilita tanto a identificação da quantidade disponível por plataforma, como a comparação entre elas. O gráfico ainda permite a visualização da quantidade exata de dados ao passar do mouse (*mouseover*) sobre cada barra.

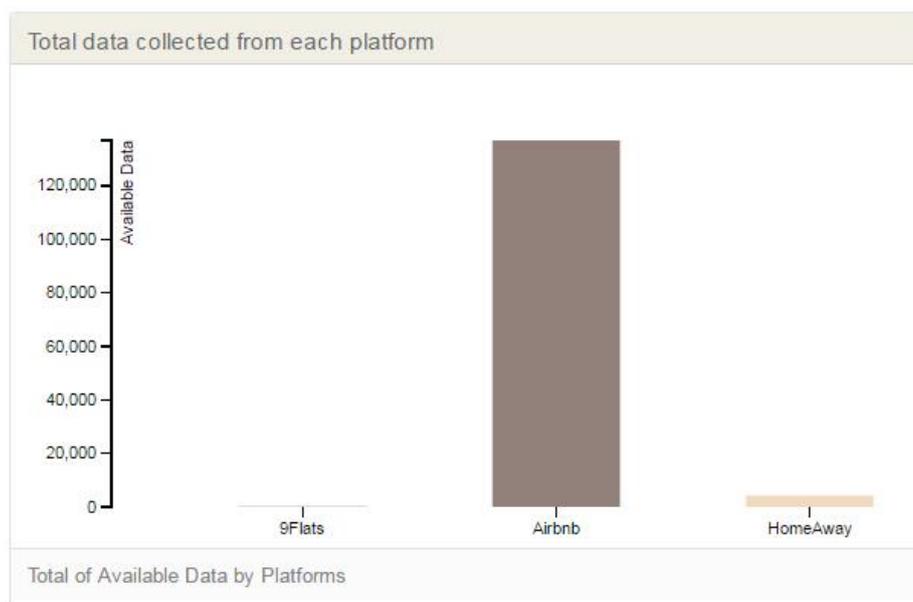


Figura 11: Total de Dados coletados por cada plataforma [Elaboração própria]

6.1.3.2 Frequência das Polaridades

Após a apresentação do total de dados disponíveis, os dois próximos gráficos também mostram dados consolidados, com ênfase nas polaridades identificadas durante a análise: os dados coletados totalizados por plataforma *versus* polaridade, em valores absolutos percentuais. Esses gráficos podem ser vistos nas figuras 14 e 15.

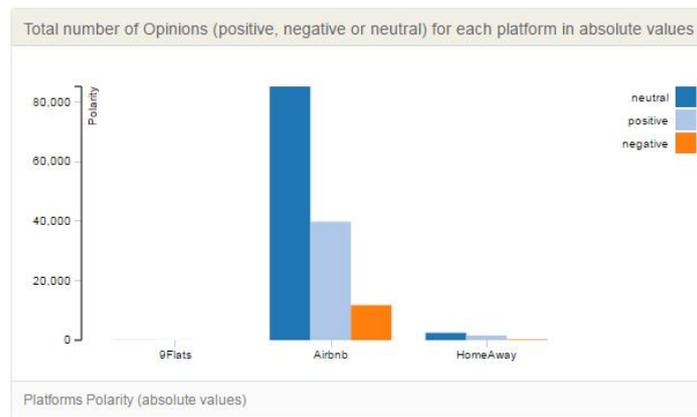


Figura 12: Quantidade de opiniões por cada plataforma em valores absolutos

[Elaboração própria]

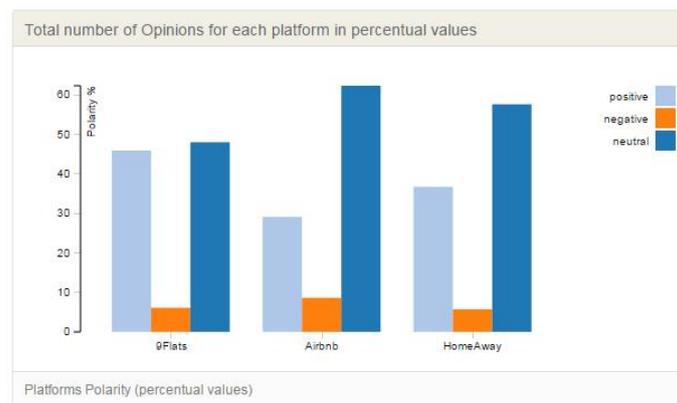


Figura 13: Quantidade de opiniões por cada plataforma em valores percentuais

[Elaboração própria]

6.1.3.3 Variação das Polaridades em uma Frequência Temporal

Após serem apresentados dados totais, que permitem uma visão mais ampla e comparativa entre as plataformas, faz-se necessário analisá-las de forma mais próxima. Para isso foi utilizado um gráfico de série temporal, onde é possível analisar a variação no dia-a-dia das polaridades de cada plataforma, de forma individual. A partir dessa visualização é possível, por exemplo, identificar se as polaridades se mantêm ao longo

dos dias, representando assim um aspecto de estabilidade na opinião ou se variam, mostrando que as opiniões sobre as mesmas ainda não estão totalmente formadas ou como as plataformas são afetadas, por exemplo, por eventos casuais. A frequência temporal de cada plataforma pode ser visualizada nas figuras 14, 15 e 16.

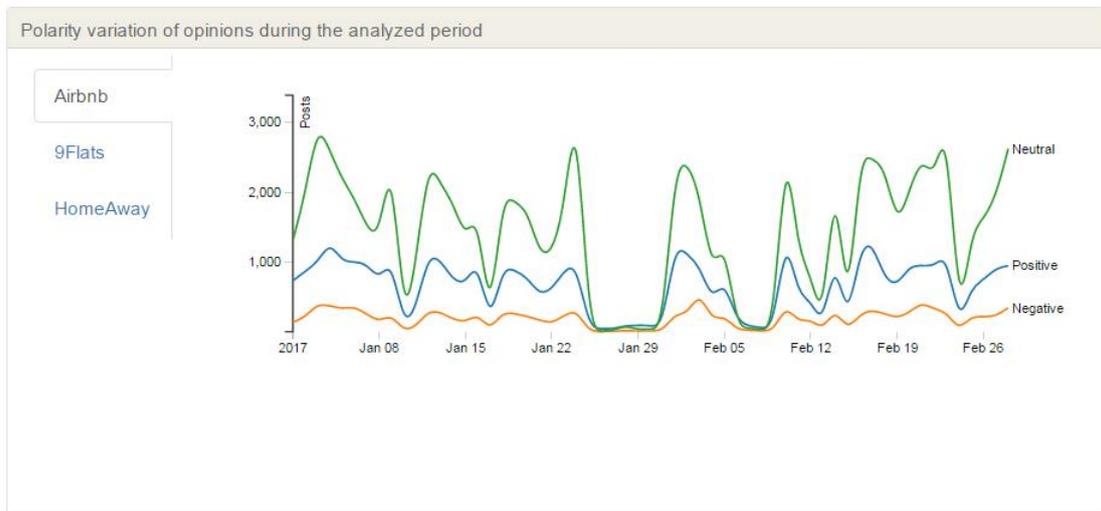


Figura 14: Frequência Temporal da Plataforma Airbnb [Elaboração própria]

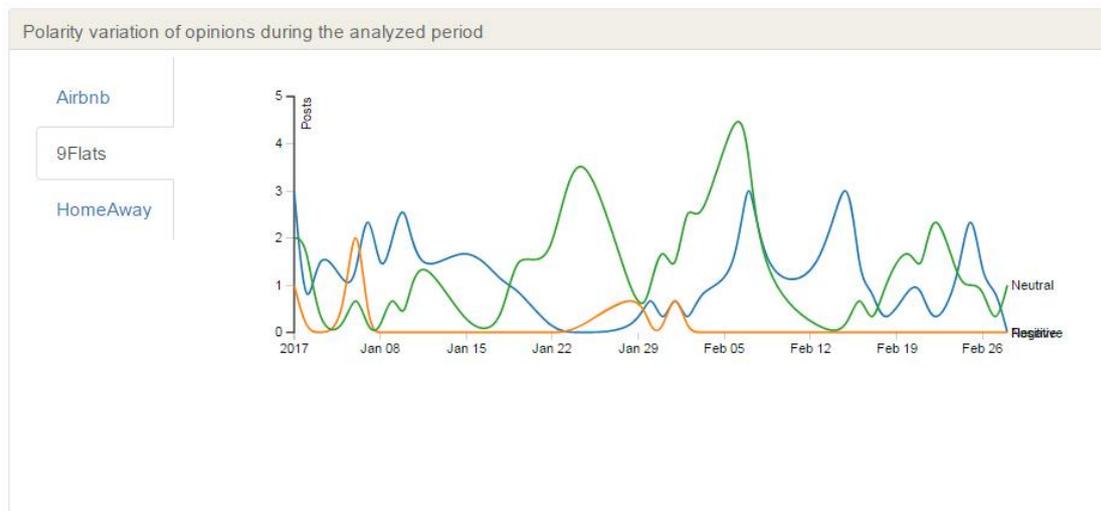


Figura 15: Frequência Temporal da Plataforma 9Flats [Elaboração própria]

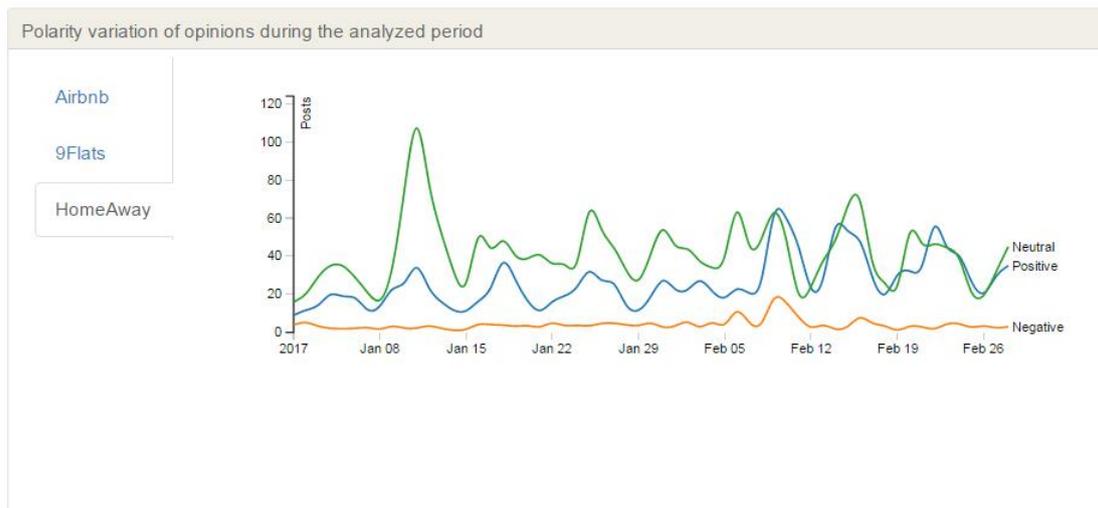


Figura 16: Frequência Temporal da Plataforma HomeAway [Elaboração própria]

6.1.3.4 Nuvem de Termos mais comentados

Esse elemento apresenta os termos mais comentados, extraídos de todos os dados coletados, e individualizados por plataformas. O processo de extração desses termos foi realizado com a utilização de um algoritmo de *POS* (do inglês *Parts-Of-Speech* — parte do discurso) *tagging*, implementado em um script escrito na linguagem python. Em linhas gerais, o *POS Tagging* pode ser definido como um processo de rotulação de elementos textuais — normalmente palavras e pontuação — que tem como finalidade evidenciar a estrutura gramatical de um determinado trecho de texto. Em reconhecimento e síntese de fala, seu uso é útil para a extração de termos, para a desambiguação, para a composição de novas frases e pesquisa lexicográfica. Logo, os termos ora extraídos e aqui apresentados não representam simplesmente as palavras mais repetidas, mas sim os termos mais importantes. O tamanho da fonte no gráfico indica a proporção da importância do termo. Logo, a escala de importância varia dos mais importantes, em tamanho de fonte maior, até os menos importantes. Também é possível, ao passar do mouse (*mouseover*) visualizar a quantidade de vezes que o termo aparece nos dados obtidos.

Posts and Polarity (if opinions are positive, negative or neutral). Texts collected from Twitter, Facebook and from the platforms themselves

Search 10

Id	Date	Platform	Comment	Polarity
9	2017-02-28	Airbnb	The apartment is really beautiful and the neighbourhood very calm. It was the perfect place for a short and calm stay.	Positive
10	2017-02-26	Airbnb	Das außergewöhnliche Innendesign und die gemütliche Einrichtung machen die Wohnung wirklich zu einem besonderen Schmuckstück. Zudem war alles tip top sauber & vollständig ausgestattet. Eine sehr zu empfehlende Unterkunft für einen unvergesslichen Wien-Trip!	Positive
11	2017-02-24	Airbnb	The Airbnb was exactly like it's shown in the pictures! Beautiful, bright and clean. It's not in the centre of town but good transport links. Only downtown would be to have directions to the house in English!	Positive
12	2017-02-20	Airbnb	Die Wohnung von Clara ist ein echtes Schmuckstück, hell, freundlich, lichtdurchflutet und in einer ruhigen Lage. Trotzdem ist die U-bahn fußläufig schnell zu erreichen. Die Wohnung war sehr sauber und mit allem nötigen ausgestattet. Durch die vielen Holzbalken und die großflächigen Fensterfronten fühlt man sich fast ein bisschen wie auf einem Schiff. Ich habe mich sehr Wohl geföhlt kann die Wohnung sehr weiterempfehlen.	Positive
13	2017-02-26	Airbnb	This place is a gem! I will definitely be visiting again in the future.	Neutral
14	2017-02-16	Airbnb	Wunderschöne Atmosphäre in Holzarchitektur. Sehr sauber und ruhig, kann ich nur weiterempfehlen!	Positive
15	2017-02-11	Airbnb	Lovely, bright apartment with awesome interior design. It is very close to U-bahn (10min) as well as to supermarket, bakery and restaurants. We would highly recommend this property!	Positive
16	2017-02-04	Airbnb	Lovely accommodation with a lovely host. Beautiful, bright apartment with Underground station just a 10 minute walk away. Supermarket, bakeries and restaurants within a 2 minute walk. Would recommend. Thank you Clara!	Positive
17	2017-01-18	Airbnb	The host canceled this reservation 37 days before arrival. This is an automated posting.	Neutral
18	2017-01-03	Airbnb	Ringraziamo Clara per l'ospitalità. L'appartamento è molto bello ed ubicato in una bella zona residenziale. Fermata metro Ober St. Veit raggiungibile a piedi in meno di 10 minuti; con una corsa di circa 15 minuti si raggiunge comodamente il centro città, senza cambi. Castello di Schonbrunn nelle vicinanze, a sole 4 fermate di metropolitana.	Positive

Showing 1 to 10 of 141295 entries

« < 1 2 3 4 5 > »

Figura 18: Grid de visualização de Opiniões, datas, e polaridades [Elaboração própria]

A partir da utilização dos elementos gráficos acima mencionados, assim como dos textos propriamente ditos, foi possível a realização de uma análise qualitativa dos resultados. Nessa análise foi possível identificar evidências da relevância das informações obtidas, sobretudo em virtude das características dos dados contidos em cada fonte de dados utilizada. No Twitter e no Facebook, por exemplo, a maioria dos dados são mais gerais, fornecendo informações sobre as próprias plataformas.

Considerando as duas fontes de dados acima, um dos objetivos da análise qualitativa foi examinar as opiniões dos usuários em relação a algumas características que são muito importantes quando se procura alugar uma acomodação. Por exemplo: como as plataformas ajudam o seu usuário na ocorrência de um problema durante a negociação do serviço? Nesse caso, como a abordagem proposta disponibiliza como insumo todos os textos coletados, uma busca pelos termos ‘no response’ e ‘inquiry’ traz à tona a experiência de vários usuários em relação ao tempo de resposta do suporte das plataformas e como elas lidam com esse tipo de problema.

Por outro lado, a análise nos dados coletados provenientes das próprias plataformas, ou seja, os *reviews* publicados por seus usuários após terem alugado uma acomodação através da plataforma, mostram informações mais específicas, direcionadas às próprias acomodações e não à plataforma em si.

6.2 Aplicação do Questionário On-line

O questionário on-line, apresentado na Seção 2.1.1.2, foi disponibilizado em uma página web, desenvolvida especificamente para esse fim. Dentro dos critérios de população e amostragem definidos, obteve-se, ao final, a participação de 22 indivíduos. Das 22 respostas obtidas, 2 foram descartadas por não conterem nenhuma informação

relevante (Figura 19). O Anexo 3 contém um extrato de todas as respostas. A avaliação, mediante os objetivos desse estudo, é descrita a seguir.

6.2.1 Análise do Resultado da Aplicação do Questionário On-line

Analisando o primeiro objetivo, descrito no início dessa Seção, pode-se concluir que uma análise simples seja necessária para a sua validação. Nesse sentido, os resultados apresentados na Seção 6.1 já seriam suficientes, uma vez que nos mesmos foram mostrados os novos dados adquiridos a partir da aplicação da abordagem proposta. Em termos meramente quantitativos, os dados adquiridos representam novos dados a serem utilizados com os existentes atualmente no Discover2Share. Por outro lado, é preciso analisar tais dados não apenas em termos de quantidade, mas, sobretudo, de qualidade, uma vez que nem sempre ter em mãos mais dados representa estar de posse de informações úteis para a tomada de decisão. Logo, o estudo de caso realizado possibilitou analisar a proposta de solução apresentada nesse trabalho, englobando os dois objetivos específicos.

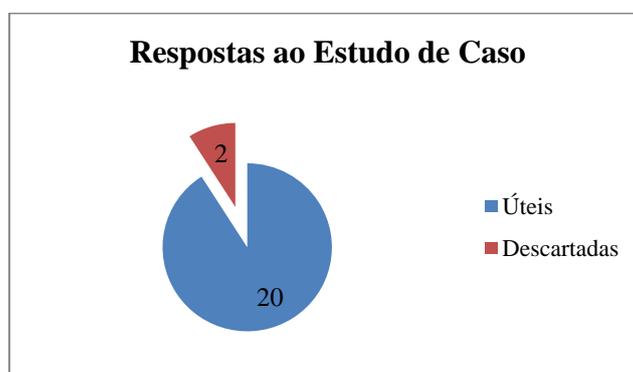


Figura 19: Quantidade de Respostas obtidas no Estudo de Caso

Algumas questões foram definidas para orientar a análise do resultado do estudo de caso:

- Os dados do primeiro conjunto (os existentes atualmente no Discover2Share) são suficientes, sozinhos, para ajudar na identificação e escolha de uma plataforma pelos usuários?

- Os usuários mudaram de opinião ao visualizarem o segundo conjunto de dados?
- Os dados do segundo conjunto (os provenientes da aplicação da abordagem apresentada nesse trabalho) fornecem mais insumos e melhoram o processo de identificação e escolha de uma plataforma pelos usuários?

Sobre a primeira questão (Figura 20), apenas cinco participantes mencionaram alguns dos atributos existentes no primeiro conjunto de dados para justificar a plataforma escolhida. Foram citados: ‘ano de fundação’; ‘possui aplicativo mobile’, ‘estrutura de lucro’, ‘duração do serviço’ e ‘meios de pagamento aceitos’.

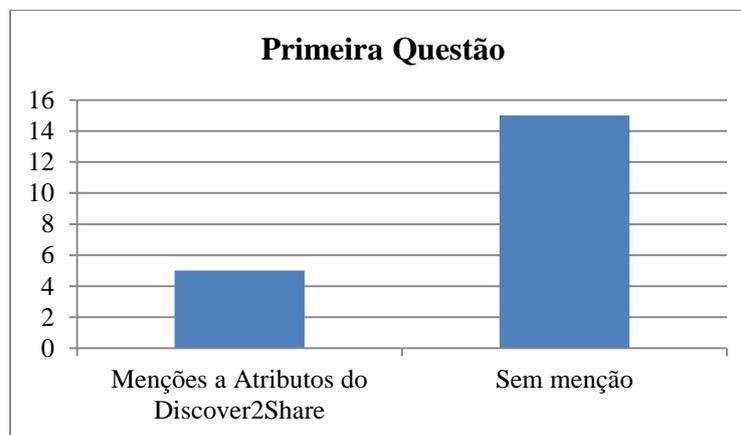


Figura 20: Menções a atributos do Discover2Share feitas na Primeira Questão

Alguns justificaram a escolha baseados em conhecimento prévio, como mencionar que a plataforma escolhida é a mais conhecida no Brasil ou que obteve algum tipo de indicação de terceiros. Outros mencionaram que escolheram qualquer uma apenas por se tratar de um campo obrigatório do questionário.

Foram obtidas também respostas sobre as quais não foi possível inferir conhecimento, comparando-as com os dados em questão. Por exemplo: ‘segurança’, ‘parece mais consistente’, ‘é mais fácil’ e ‘é indiferente, já que é necessário testar a plataforma’.

Por fim, foram coletadas algumas respostas diretas apontando que o conjunto de dados não fornece nenhum insumo para a identificação e escolha da plataforma. Por exemplo: ‘os dados apresentados não me dizem muita coisa. Parecem dados técnicos..’; ‘.. é impossível analisar os dados apresentados..’; ‘.. não consegui entender o significado desses dados...’; ‘.. as informações apresentadas de cada plataforma não faz(sic) diferença neste caso..’; ‘.. eles são de alguma forma semelhantes..’ e ‘..Airbnb e 9Flats tem descrições muito semelhantes..’.

A análise das respostas acima aponta que **a resposta para a primeira questão é que o primeiro conjunto de dados (os existentes atualmente no Discover2Share) não é suficiente, sozinho, para ajudar na identificação e escolha de uma plataforma.**

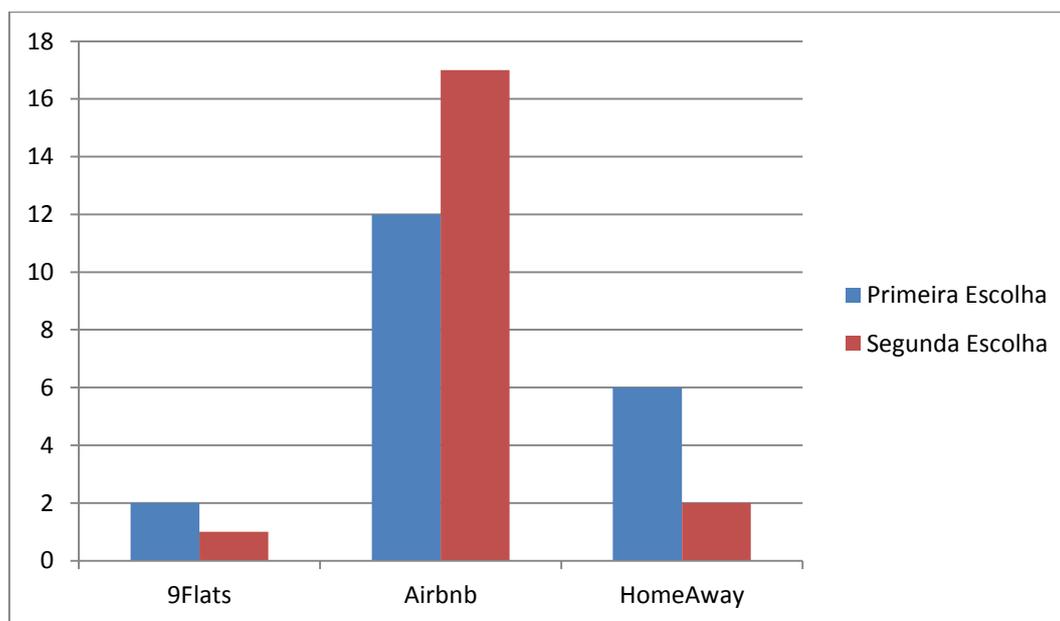


Figura 21: Mudança na Plataforma escolhida entre o primeiro e o segundo conjunto de dados

Dando prosseguimento, os dados coletados foram analisados tendo em vista a segunda questão. Das 20 respostas, houve mudança na escolha da plataforma em seis casos (Figura 21). Dentro desses seis casos, a escolha na primeira opção foi baseada nos atributos do primeiro conjunto de dados em três casos. Nos demais casos, a escolha foi

pela obrigatoriedade de preenchimento. **Com base nessas respostas, e também na análise da primeira questão, é perceptível, ainda que de forma tímida, a influência positiva do segundo conjunto de dados como fator decisório.**

Após a análise sob a perspectiva das duas primeiras questões, os dados foram analisados em virtude da terceira questão, ou seja, verificar se os dados do segundo conjunto melhoram ou não o processo de identificação e escolha da plataforma. Das 20 respostas, apenas três não citam diretamente os dados contidos no segundo conjunto para justificar a sua escolha. Nesses três casos os comentários foram inconclusivos sobre a interferência dos dados: ‘por que é a melhor’, ‘mais familiar’ e ‘mais confortável para mim’.

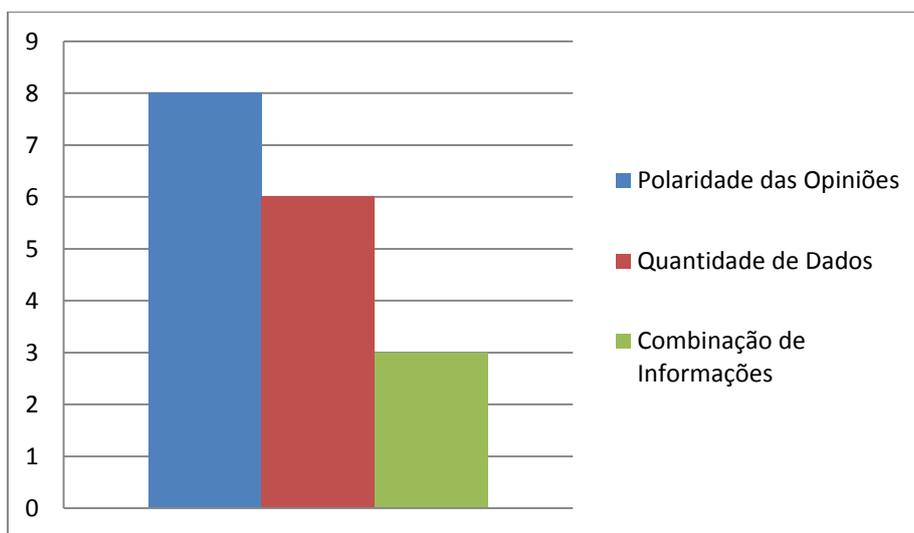


Figura 22: Fatores citados como relevantes para a tomada de decisão

Nos demais 17 (Figura 22), os dados disponibilizados são expressamente citados como fator de decisão. Em oito situações é citada a polaridade das opiniões. Em seis é citada a quantidade de dados. Em três são mencionadas fatores que demonstram uma análise mais aprofundada sobre os dados pelos indivíduos. Em tais casos são citados ‘a maior atividade dos usuários da plataforma’, ‘... parece melhor de acordo com as opiniões lidas...’ e ‘HomeAway parece oferecer um serviço melhor, mas tem um marketing menor do que a Airbnb’ . **Frente ao exposto acima, conclui-se, em relação**

à terceira questão, que os dados disponibilizados através da abordagem aplicada nesse trabalho tanto fornecem mais insumos quanto melhoram o processo de identificação e escolha de uma plataforma.

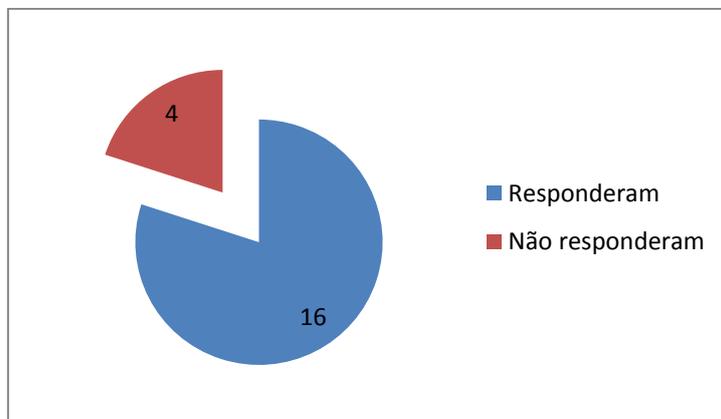


Figura 23: Quantidade de respostas obtidas sobre a validade do estudo de caso

Por fim, a aplicação do questionário trouxe ainda outras informações relevantes, que tanto corroboraram com as conclusões acima como oferecem outros pontos de vista sobre o experimento. Tais informações foram obtidas da última pergunta do questionário (“o que você achou desse experimento”). Dos 20 participantes (Figura 23), apenas quatro não responderam a essa questão. Dos 16 respondentes, três emitiram comentários muito simples e que, por isso, não foram analisados: ‘bom’, ‘insuficiente’ e ‘realmente incrível’. Nas demais respostas foram observados comentários elogiando o experimento e ressaltando a validade do uso de opiniões para a tomada de decisão. Também foram encontradas críticas, com comentários taxando o estudo de caso como confuso ou mencionando que a forma como as informações foram dispostas talvez não fosse a mais propícia para o que se pretendia.

Além das informações registradas no questionário e armazenadas, também foram coletadas opiniões sobre o experimento em conversas informais, tanto com indivíduos que participaram da pesquisa, como com outros que não participaram. Nessas conversas

os pontos mais importantes discutidos estavam relacionados à exibição dos dados, mais precisamente à importância de ter sido possível visualizar os mesmos em dispositivos móveis. Esse ponto, não visto em discussão em outros trabalhos relacionados ou na referência bibliográfica utilizada, desponta como importante para ser tratado mais a fundo em trabalhos futuros.

7. Conclusão

Nesse Capítulo é apresentada a conclusão dessa dissertação. Além disso, alguns trabalhos futuros são delineados.

Esse trabalho abordou a importância das plataformas ponto-a-ponto (P2P) de consumo e compartilhamento colaborativo (SCC) como canal para o compartilhamento de recursos entre indivíduos. Tais plataformas têm se expandido rapidamente, facilitando a vida dos indivíduos e promovendo ainda os conceitos relacionados à economia compartilhada. Esse crescimento, além de benefícios, gera ainda alguns problemas, sobretudo pela quantidade e diversidade de plataformas existentes. Vindo ao encontro desses problemas algumas soluções também têm sido propostas, entre elas o diretório de busca e descoberta de plataformas Discover2Share.

Ao ser analisado, e também comparado às propostas similares existentes, o Discover2Share apresenta significativos progressos. Entretanto, o mesmo possui um importante gap: embora disponibilize informações estruturadas sobre várias plataformas, esses dados tendem a não ser suficientes para permitir uma identificação e escolha mais precisas pelos usuários do diretório.

A motivação para esse trabalho foi aproveitar os progressos alcançados pelo Discover2Share e adicionar conteúdo – sobretudo opiniões – a partir de conteúdo social (posts publicados pelos usuários em redes sociais e nas próprias plataformas) que possa ser usado como contexto externo, juntamente com os dados existentes atualmente, para melhorar o processo de identificação de plataformas de consumo e compartilhamento colaborativo.

7.1 Contribuições

Ao longo dos Capítulos e Seções anteriores, a proposta de solução e sua metodologia foram detalhadas e, a seguir, aplicadas em dois Estudos de Casos, um experimento que coletou e analisou os dados relativos a três plataformas que fornecem recursos do tipo ‘acomodações’: 9Flats, Airbnb e HomeAway; e outro através de um questionário online, onde a proposta de solução foi avaliada junto a alguns usuários. Seguindo o objetivo principal desse trabalho, que era enriquecer o Discover2Share e melhorar o processo de identificação de plataformas por seus usuários, o resultado dos experimentos demonstrou a validade da proposta, comprovada através dos estudos de caso aplicados.

Com isso, a primeira contribuição desse trabalho foi o método criado, que mostrou ser capaz de obter novos e valiosos dados a partir do conteúdo social. Outra contribuição, ainda em relação ao método, diz respeito ao seu caráter multidisciplinar, ou seja, o mesmo pode ser aplicado a diferentes plataformas, que forneçam diferentes tipos de recursos e até mesmo em outros domínios.

No campo da pesquisa científica, esse trabalho contribuiu ao fazer uso de um estudo de caso que comprovou a importância da mineração de opiniões nos processos intensivos em conhecimento de escolha por produtos ou serviços. Além disso, apresentou indícios da validade da utilização de textos traduzidos na mineração de opiniões. Outra contribuição importante no campo da pesquisa científica, e pouco vista em outros trabalhos, foi a utilização de múltiplas fontes de dados no campo da mineração de opiniões.

7.2 Trabalhos Futuros

Para continuidade e aprofundamento dos estudos aqui abordados são sugeridos três trabalhos futuros: o primeiro diz respeito a avaliar a viabilidade de aplicação da abordagem aqui adotada em tempo real. Nesse sentido, algumas adaptações, sobretudo na coleta dos dados e, conseqüentemente, na sua quantidade, precisariam ser endereçadas e validadas a fim de se confirmar se os seus resultados equivaleriam aos apresentados nesse trabalho. O segundo trabalho futuro sugerido diz respeito à validação dos algoritmos de análise de sentimento utilizados, e que fazem parte da ferramenta iFeel, sobretudo no que diz respeito às suas particularidades versus as fontes de dados utilizadas. Indícios apontam que uma abordagem utilizando pesos na definição da polaridade preponderante poderia apresentar resultados mais precisos, uma vez que alguns algoritmos tem melhor desempenho do que outros em virtude dos tipos/fontes de dados utilizadas. Um terceiro trabalho sugerido é a avaliação aprofundada do método multi-idioma.

Referências Bibliográficas

- [1] Gansky, L.: *The Mesh: Why the Future of Business is Sharing*. New York, Penguin (2010).
- [2] Botsman, R., Rogers, R.: *O que é meu é seu: Como o Consumo Colaborativo vai Mudar o Nosso Mundo*. Porto Alegre, Bookman (2009).
- [3] Dubois, E., Schor, J., Carfagna, L.: *Connected Consumption: A Sharingeconomy Takes Hold*. Rotman Management, pp. 50–55 (2014).
- [4] Hoffen, M., Matzner, M., Chasin, F.: *Designing an Ontology-based Web Directory for the Discovery of Sharing and Collaborative Consumption Platforms*. University of Münster, Muenster, Germany (2015).
- [5] Maes, P., Sharnand, U.: *Social Information Filtering: Algorithms for Automating “word of mouth”*. In: *Human Factors in Computing Systems*, pp. 210-217 (1995).
- [6] Zhao, K., Stylianou, A., Zheng, Y.: *Sources and Impacts of Social Influence from Online Anonymous User Reviews*. In: *Journal Information & Management* (2017).
- [7] Berger, J., Schwartz, E.: *What Drives Immediate and Ongoing Word of Mouth?* In: *Journal of Marketing Research*, v. 48(5), pp. 869–880 (2011).
- [8] Gruen, T., Osmonbekov, T., Czaplewski, A.: *Customer-to-Customer Exchange: Its MOA Antecedents and its Impact on Value Creation and Loyalty*. In: *Journal of the Academy of Marketing Science*, v. 35(4), pp. 537–549 (2007).
- [9] Smith, D., Menon, S., Sivakumar, K.: *Online Peer and Editorial Recommendations, Trust, and Choice in Virtual Markets*. In: *Journal of Interactive Marketing*, v. 19(3), pp.15–37 (2005).
- [10] Murray, K.: *A Test of Services Marketing Theory: Consumer Information Acquisition Activities*. In: *The Journal of Marketing*, pp. 10–25 (1991).

- [11] Bickart, B., Schindler, R.: Internet Forums as Influential Sources of Consumer Information. In: *Journal of Interactive Marketing*, v. 15(3), pp. 31–40 (2001).
- [12] Sweeney, J., Soutar, G., Mazzarol, T.: Word of Mouth: Measuring the Power of Individual Messages. In: *European Journal of Marketing*, v. 46, pp. 1–37 (2011).
- [13] Bei, L., Chen, E., Widdows, R.: Consumers' Online Information Search Behavior and the Phenomenon of Search vs. Experience Products. In: *Journal of Family and Economic Issues*, v. 25(4), pp. 449–467 (2004).
- [14] Dellarocas, C.: The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. In: *Management Science*, v. 49(10), pp. 1407–1424 (2003).
- [15] Childers, T., Carr, C., Peck, J., Carson, S.: Hedonic and Utilitarian Motivations for Online Retail Shopping Behavior. In: *Journal of Retailing*, v. 77(4), pp. 511–535 (2002).
- [16] Mathwick, C., Wiertz, C., De Ruyter, K.: Social Capital Production in a Virtual P3 Community. In: *Journal of Consumer Research*, v. 34(6), pp. 832–849 (2008).
- [17] Singh, J., Irani, S., Rana, P., Dwivedi, K., Saumya, S., Roy, K.: Predicting the “Helpfulness” of Online Consumer Reviews. In: *Journal of Business Research*, v. 70, pp. 346–355 (2017).
- [18] Sotiriadis, M., Van Zyl, C.: Electronic Word-Of-Mouth and Online Reviews in Tourism Services: The Use of Twitter by Tourists. In: *Electronic Commerce Research*, v. 13(1), pp. 103 (2013).
- [19] Chen, J., Teng, L., Yu, Y., Yu, X.: The Effect of Online Information Sources on Purchase Intentions Between Consumers with High and Low Susceptibility to Informational Influence. In: *Journal of Business Research*, v. 69(2), pp. 467–475 (2016).

- [20] Botsman, R.: Sharing's not just for Start-Ups. In: *Harv Bus*, v. 92(9), pp. 23–26 (2014).
- [21] Hamari, J., Sjöklint, M., Ukkonen, A.: The Sharing Economy: Why People Participate in Collaborative Consumption. *J Assn Inf Sci Tec*, v. 67, pp. 2047–2059 (2016).
- [22] Lessig L.: *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. Penguin, New York (2008).
- [23] Calder, B., Malthouse, E.: Managing Media and Advertising Change with Integrated Marketing. In: *Journal of Advertising Research*, vol. 45, no. 04, pp. 356–361 (2005).
- [24] Bardhi , F., Eckhardt, G.: Access-Based Consumption: The Case of Car Sharing. In: *Journal of Consumer Research*, vol. 39, no. 4, pp. 881–898 (2012).
- [25] Bendapudi , N., Leone, R.: Psychological Implications of Customer Participation in Co-Production. In: *Journal of Marketing*, vol. 67, no. 1, pp. 14–28 (2013).
- [26] Mills , P., Morris, J.: Clients as "Partial" Employees of Service Organizations: Role Development in Client Participation. In: *Academy of Management Review*, vol. 11, no. 4, pp. 726–735, 1986. [Online]. Available: <http://amr.aom.org/content/11/4/726.short>
- [27] Kreyer , N., Pousttchi, K., Turowski, K.: Standardized Payment Procedures as Key Enabling Factor for Mobile Commerce. In *Proceedings of the Third International Conference on E-Commerce and Web Technologies (ECWeb 2002)*, ser. *Lecture Notes in Computer Science (LNCS)*, K. Bauknecht, A. M. Tjoa, and G. Quirchmayr, Eds. Aix-en-Provence, France: Springer, 2007, vol. 2455, pp. 400–409.

- [28] Atkinson, J., Salas, G., Figueroa, A.: Improving Opinion Retrieval in Social Media by Combining Featuresbased Coreferencing and Memory-Based Learning. In: Information Sciences, v. 299, pp. 20-31 (2015).
- [29] Agarwal, B., Mittal, N., Bansal, P., Garg, S.: Sentiment Analysis Using Common-Sense and Context Information. In: Computational intelligence and neuroscience, v. 2015, pp. 30 (2015).
- [30] Wiebe, J.: NRRC Summer Study Jan Wiebe and Group (University of Pittsburgh) on 'Subjective' Statements. (2002).
- [31] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs Up? Sentiment Classification Using Machine Learning Techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing, v. 10, pp. 79-86 (2002).
- [32] Turney, P.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417-424 (2002).
- [33] Dashtipour, K., Poria, S., Hussain, A. et al.: Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques. In: Cogn Comput, vol. 8, pp. 757-771 (2016).
- [34] Araújo, M., Reis, J., Pereira, A., Benvenuto, F. : An Evaluation of Machine Translation is Multilingual Sentence-Level Analysis Sentiment. In: Proceedings of the 31st Annual ACM Symposium on Applied Computing. ACM (2016).
- [35] Morinaga, S., Yamanishi, K., Tateishi, K., Fukushima, T.: Mining Product Reputations on The Web. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 341-3492 (2002).

- [36] Gamon, M., Aue, A., Corston-Oliver, S., Ringger, E.: Pulse: Mining Customer Opinions from Free Text. In: international symposium on intelligent data analysis, pp. 121-132 (2005).
- [37] Chen, C., Ibekwe-SanJuan, F., SanJuan, E., Weaver, C.: Visual Analysis Of Conflicting Opinions. In: Visual Analytics Science and Technology, IEEE Symposium On, pp. 59-66 (2006).
- [38] Draper, G., Riesenfeld, R.: Who Votes for What? A Visual Query Language for Opinion Data. *IEEE transactions on visualization and computer graphics*, v. 14(6), pp. 1197-1204 (2008).
- [39] Liu, B., Hu, M., Cheng, J.: Opinion Observer: Analyzing and Comparing Opinions on the Web. In: Proceedings of the 14th international conference on World Wide Web, pp. 342-351 (2005).
- [40] Oelke, D., Hao, M., Rohrdantz, C., Keim, D. A., Dayal, U., Haug, L. E., Janetzko, H.: Visual Opinion Analysis of Customer Feedback Data. In: Visual Analytics Science and Technology, IEEE Symposium on, pp. 187-194 (2009).
- [41] Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., Qu, H.: Opinionseer: Interactive Visualization of Hotel Customer Feedback. *IEEE transactions on visualization and computer graphics*, v. 16(6), pp. 1109-1118 (2010).
- [42] Araújo, M., Gonçalves, P., Cha, M., Benevenuto, F.: iFeel: A Web System that Compares and Combines Sentiment Analysis Methods. In: Proceedings of the 23rd International Conference on World Wide Web Seoul, Korea, pp. 75-78 (2014).
- [43] Nielsen, F.: A new ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. *arXiv preprint arXiv:1103.2903* (2011).
- [44] Mohammad, S., Turney, P.: Crowdsourcing a Word–Emotion Association Lexicon. In: *Computational Intelligence*, v. 29(3), pp. 436-465 (2013).

- [45] Hannak, A., Anderson, E., Barrett, L., Lehmann, S., Mislove, A., Riedewald, M.: Tweetin'in the Rain: Exploring Societal-Scale Effects of Weather on Mood. In ICWSM (2012).
- [46] Mohammad, S.: #Emotional Tweets. In: Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (pp. 246-255). Association for Computational Linguistics (2012).
- [47] Hu, M., Liu, B.: Mining and Summarizing Customer Reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining pp. 168-177, (2004).
- [48] Mohammad, S., Kiritchenko, S., Zhu, X.: NRC-Canada: Building the State-Of-The-Art in Sentiment Analysis of Tweets. arXiv preprint arXiv:1308.6242 (2013).
- [49] Levallois, C.: Umigon: Sentiment Analysis for Tweets Based on Lexicons and Heuristics. Atlanta, Georgia, USA, 414, (2013).
- [50] Hutto, C., Gilbert, E.: Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In: Eighth international AAI conference on weblogs and social media (2014).
- [51] Gonçalves, P., Benevenuto, F., Cha, M.: PANAS-t: A Psychometric Scale for Measuring Sentiments on Twitter. In: arXiv preprint arXiv:1308.1857 (2013).
- [52] Dodds, P., Danforth, C.: Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, And Presidents. In: Journal of Happiness Studies, v. 11(4), pp. 441-456 (2009).
- [53] Cambria, E., Speer, R., Havasi, C., Hussain, A.: Senticnet: A Publicly Available Semantic Resource for Opinion Mining. In: AAI Fall Symposium Series (2010).

- [54] Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment Strength Detection for the Social Web. In: Journal of the American Society for Information Science and Technology, v. 63(1), pp. 163-173 (2012).
- [55] Richter-von Hagen, C., Ratz, D., Povalej, R.: Towards Self-Organizing Knowledge Intensive Processes. In: Journal of Universal Knowledge Management, v. 2, pp. 148-169 (2005).
- [56] Maldonado, M.: Análise do Impacto das Políticas de Criação e Transferência de Conhecimento em Processos Intensivos em Conhecimento: Um Modelo de Dinâmica de Sistemas. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento). Programa de Pós-graduação em Engenharia e Gestão de Conhecimento. Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brasil (2008).
- [57] Matta, F.: Pesquisa de Marketing: Metodologia, Planejamento. São Paulo: Atlas (2001).
- [58] Kirk, J., Miller, L.: Reliability and Validity in Qualitative Research. Sage (1986).
- [59] Bromley, D.: The Case-Study Method in Psychology and Related Disciplines. John Wiley & Sons (1986).
- [60] Yin, R.: Case Study Research (1994).
- [61] Malhotra, N.: Marketing Research: An Applied Orientation. 5/e, Pearson Education India (2008).
- [62] Patton, M.: Qualitative Evaluation and Research Methods. SAGE Publications, inc (1990).
- [63] Babbie, E.: The Practice of Social Research. Nelson Education (2015).

ANEXO 1

Case Study

Improving the Decision Making Process of choosing a Sharing-Economy Platform through Opinions as External Context.
Consider you have 3 options of accommodation platforms (9Flats, Airbnb and HomeAway) and you must choose one.

Which data would be more relevant to guide you in your decision?

1st Option: data describing features of the platform or
2nd Option: social contents as opinions about the platforms.

*The response time of the questions is approximately 2 minutes.

1st Option

Airbnb		9Flats		HomeAway	
Predicate	Data	Predicate	Data	Predicate	Data
type	Platform	type	Platform	type	Platform
type	NamedIndividual	type	NamedIndividual	type	NamedIndividual
foundationPlace	san-francisco-ca	foundationPlace	deu	foundationPlace	usa
foundationPlace	usa	foundationPlace	hamburg	foundationPlace	austin-tx
residence	san-francisco	residence	deu	residence	usa
residence	usa	residence	hamburg	residence	austin-tx
acceptePaymentMethods	per-transaction	acceptePaymentMethods	per-transaction	acceptePaymentMethods	per-transaction
acceptePaymentMethods	accommodations	acceptePaymentMethods	accommodations	acceptePaymentMethods	membership-fee
consumeInvolvement	in-between	consumeInvolvement	in-between	consumeInvolvement	consumation
consumeInvolvement	in-between	consumeInvolvement	in-between	consumeInvolvement	accommodations
durationOfServiceMax	days	durationOfServiceMax	weeks	durationOfServiceMax	in-between
durationOfServiceMax	days	durationOfServiceMax	weeks	durationOfServiceMax	weeks
grantsToBeResourceProvider	private	grantsToBeResourceProvider	private	grantsToBeResourceProvider	days
grantsToBeResourceProvider	private	grantsToBeResourceProvider	private	grantsToBeResourceProvider	private-and-business
hasAccessType	integrated	hasAccessType	integrated	hasAccessType	integrated
hasMobileApplicationSupport	mobile-application-android-airbnb	hasMobileApplicationSupport	mobile-application-android-airbnb	hasMobileApplicationSupport	mobile-application-android-homeaway
hasMobileApplicationSupport	mobile-application-ios-airbnb	hasMobileApplicationSupport	mobile-application-ios-airbnb	hasMobileApplicationSupport	mobile-application-ios-homeaway
hasMobileApplicationSupport	mobile-application-ios-airbnb	hasMobileApplicationSupport	mobile-application-ios-airbnb	hasMobileApplicationSupport	mobile-application-ios-homeaway
moneyFlow	c2b2c	moneyFlow	c2b2c	moneyFlow	c2b2c
moneyFlow	c2b2c	moneyFlow	c2b2c	moneyFlow	c2b2c
motivesForConsumption	economical	motivesForConsumption	economical	motivesForConsumption	economical
motivesForConsumption	economical	motivesForConsumption	economical	motivesForConsumption	social
motivesForConsumption	social	motivesForConsumption	social	motivesForConsumption	social
necessaryTimeForTransaction	deferred	necessaryTimeForTransaction	deferred	necessaryTimeForTransaction	deferred
necessaryTimeForTransaction	deferred	necessaryTimeForTransaction	deferred	necessaryTimeForTransaction	deferred
profitStructure	profit-from-both-peer-consumers-and-peer-providers	profitStructure	profit-from-peer-providers	profitStructure	profit-from-peer-providers
profitStructure	profit-from-both-peer-consumers-and-peer-providers	profitStructure	profit-from-peer-providers	profitStructure	profit-from-peer-providers
serveArea	global	serveArea	global	serveArea	global
serveArea	global	serveArea	global	serveArea	global
typeOfAccessedObject	mixed	typeOfAccessedObject	mixed	typeOfAccessedObject	mixed
typeOfAccessedObject	mixed	typeOfAccessedObject	mixed	typeOfAccessedObject	mixed
foundingYear	2008	foundingYear	2011	foundingYear	2005
foundingYear	2008	foundingYear	2011	foundingYear	2006
idNumber	100	idNumber	777	idNumber	258
idNumber	100	idNumber	777	idNumber	335
idNumber	100	idNumber	777	idNumber	335
title	airbnb	title	9Flats	title	HomeAway
title	airbnb	title	9Flats	title	HomeAway
url	http://www.airbnb.com	url	http://www.9flats.com	url	http://www.homeaway.co.uk
url	http://www.airbnb.com	url	http://www.9flats.com	url	http://www.homeaway.com
identifier	100	identifier	777	identifier	258
identifier	100	identifier	777	identifier	335
identifier	100	identifier	777	identifier	335

Airbnb Details

9Flats Details

HomeAway Details

Which platform do you choose?

9Flats Airbnb HomeAway

Why?

ANEXO 2

Case Study

Improving the Decision Making Process of choosing a Sharing-Economy Platform through Opinions as External Context.
 Consider you have 3 options of accommodation platforms (9Flats, Airbnb and HomeAway) and you must choose one.

Which data would be more relevant to guide you in your decision?

- 1st Option: data describing features of the platform or
- 2nd Option: social contents as opinions about the platforms.

*The response time of the questions is approximately 2 minutes.

2nd Option

[Back to First Option](#)

Total data collected from each platform

Total of Available Data by Platforms

Total number of Opinions (positive, negative or neutral) for each platform in absolute values

Platforms Polarity (absolute values)

Total number of Opinions for each platform in percentual values

Platforms Polarity (percentual values)

Polarity variation of opinions during the analyzed period

9Flats: Most commented terms

9Flats: Most commented terms

HomeAway: Most commented terms

There were 468 terms tagged 'vacation'

Posts and Polarity (if opinions are positive, negative or neutral) - Texts collected from Twitter, Facebook and from the platforms themselves

Id	Date	Platform	Comment	Polarity
9	2017-02-28	Airbnb	The apartment is really beautiful and the neighbourhood very calm. It was the perfect place for a short and calm stay.	Positive
10	2017-02-26	Airbnb	Das außergewöhnliche Innendesign und die gemütliche Einrichtung machen die Wohnung wirklich zu einem besonderen Schmuckstück. Zudem war alles tip top sauber & vollständig ausgestattet. Eine sehr zu empfehlende Unterkunft für einen unvergesslichen Wien-Trip!!	Positive
11	2017-02-24	Airbnb	The Airbnb was exactly like it's shown in the pictures! Beautiful, bright and clean. It's not in the centre of town but good transport links. Only downfall would be to have directions to the house in English!	Positive
12	2017-02-20	Airbnb	Die Wohnung von Clara ist ein echtes Schmuckstück, hell, freundlich, lichtdurchflutet und in einer ruhigen Lage, trotzdem ist die U-Bahn fußläufig schnell zu erreichen. Die Wohnung war sehr sauber und mit allem nötigen ausgestattet. Durch die vielen Holzbalken und die großflächigen Fensterfronten fühlt man sich fast ein bisschen wie auf einem Schiff. Ich habe mich sehr Wohl gefühlt kann die Wohnung sehr weiterempfehlen.	Positive
13	2017-02-26	Airbnb	This place is a gem I will definitely be visiting again in the future.	Neutral
14	2017-02-16	Airbnb	Wunderschöne Atmosphäre in Holzarchitektur. Sehr sauber und ruhig, kann ich nur weiterempfehlen!	Positive
15	2017-02-11	Airbnb	Lovely, bright apartment with awesome interior design. It is very close to U-bahn (10min) as well as to supermarket, bakery and restaurants. We would highly recommend this property!	Positive
16	2017-02-04	Airbnb	Lovely accommodation with a lovely host. Beautiful, bright apartment with Underground station just a 10 minute walk away. Supermarket, bakeries and restaurants within a 2 minute walk. Would recommend. Thank you Clara!	Positive
17	2017-01-18	Airbnb	The host canceled this reservation 37 days before arrival. This is an automated posting.	Neutral
18	2017-01-03	Airbnb	Ringraziamo Clara per l'ospitalità. L'appartamento è molto bello ed ubicato in una bella zona residenziale. Fermata metro Ober St. Veit raggiungibile a piedi in meno di 10 minuti, con una corsa di circa 15 minuti si raggiunge comodamente il centro città, senza carro. Castello di Schönbrunn nelle vicinanze, a soli 4 fermate di metropolitana.	Positive

Showing 1 to 10 of 141295 entries

Which platform do you choose?

9Flats Airbnb HomeAway

Why?

What did you think about this experiment?

ANEXO 3

Id	Date	IP	1st Option	(1st Option) Why did you choose?	2nd Option	(2nd Option) Why did you choose?	What did you think about this experiment?
5	08/06/2017	200.222.105.250	homeaway	Os dados apresentados não me dizem muita coisa. Parecem dados técnicos. Se fosse para decidir, pensando na plataforma que já ouvi falar, seria airbnb. MAS, se obrigatoriamente, a escolha fosse baseada tão-somente nestes dados, escolheria HomeAway pelo ano de fundação.	airbnb	Devido a quantidade de informações coletadas, baseadas nas experiências de outros usuários, eu escolheria o AirBnb, pois proporcionalmente as avaliações são boas [somando neutras e positivas].	Achei pertinente, levando em conta que as opiniões de outras pessoas - que viveram a experiência da compra nessas plataformas - nos oferecem a segurança necessária para prosseguir ou não na finalização de uma reserva.
6	08/06/2017	200.201.128.133	airbnb	most used	airbnb	has a higher number of posts	good
7	08/06/2017	200.156.27.28	airbnb	It's the only one I know. It's impossible to analyse the data provided	airbnb	Most of data relate to Airbnb	
8	08/06/2017	200.156.27.28	9flats	Actually I do not know, because I do not understand the meaning of this data. I only chose "9Flats" because it is a required answer.	airbnb	It seemed more popular, and positively evaluated.	the way data was presented (text+tables in the 1st option, graphics+colors in the 2nd option) influenced my choice.
9	08/06/2017	187.13.240.38	homeaway	Well because yes.	9flats	Well, because yes	
10	08/06/2017	200.156.26.34	homeaway	Tive dificuldades para comparar os modelos, seria interessante apontar as diferenças ao apresentar cada um e talvez apontar o que seria relevante para fazer a comparação. Escolhi uma das opções ao acaso, pois não tenho conhecimento para indicar o melhor.	airbnb	Por ter mais dados coletados	Não entendi sobre o contexto, na primeira tela é difícil de comparar as tabelas (talvez um destaque no que havia de diferente ajudasse), na segunda tela teve mais informações uma forma melhor de comparar, porém diferença grande de dados coletados o que pode tender a uma escolha.
11	08/06/2017	201.51.93.151	airbnb	apenas por que é a mais conhecida, as informações apresentadas de cada plataforma não faz diferença neste caso. Já ouvi falar da airbnb e 9Flats.	airbnb	opiniões positivas	eu acho que a segunda parte da pesquisa contém dados que podem formular uma opinião com base então a escolha é influenciada por algo concreto (opiniões de outras pessoas); a primeira parte da pesquisa não influencia em nada; talvez com empresas/produtos desconhecidos aqueles dados poderiam fazer alguma diferença.
12	09/06/2017	189.31.119.152	homeaway	HomeAway seems more flexible in comparison to Airbnb and 9Flats	airbnb	Airbnb, because it has much more user activity.	
13	09/06/2017	187.126.219.182	airbnb	Because it is the most famous in Brazil	airbnb	Because it seems the best one	Could be a good experience and cheaper than a hotel
14	09/06/2017	201.94.167.150	airbnb	Segurança	airbnb	Mais familiar	Possibilidade de conhecer pessoas do local que podem indicar os melhores pontos turísticos.
15	09/06/2017	177.140.136.41	airbnb	They are somehow similar but I would choose AirBnB, the profit Structure has what seems to be an advantage in being able to profit from peer consumers has well and has mobile application in Windows. Although its not a big slice in the mobile market in case of future development has more potencial.	airbnb	Would choose Airbnb by the number of positive opinions, inspite of in terms of percentage they have more neutral comments, in absolute values we are able to see that a lot of individuals use their services, and they are seemingly satisfied.	Very interesting in terms of the 2nd option being able to view values of opinions in percentage and in global values helps to understand the acceptance of such services.
16	09/06/2017	200.222.105.250	airbnb	Das três plataformas, é a que me passa mais segurança e confiabilidade.	airbnb	O Airbnb disponibiliza uma quantidade superior de comentários, passando mais segurança e confiabilidade. Esses itens, ao meu ver, são essenciais para o usuário, em especial um usuário iniciante, que não possui muita experiência com turismo online.	Confuso, ao menos a princípio. Requer um pouco de paciência e foco para entender os dados e as informações que são pedidas.
17	09/06/2017	177.86.158.56	airbnb	Looks more consistent	airbnb	Better (and more) feedback from customers	
18	09/06/2017	146.134.48.80	airbnb	foundingYear	9flats	Platforms Polarity (percentual values)	Insufficient
19	10/06/2017	189.106.60.64	airbnb	Recommendation of other friends	airbnb	Most people recommended the platform. And the data show it is better than 9Flats and HomeAway	This study case is so confused. It is better introduce the user about what you want the user do, show the data e after make the asks.
20	10/06/2017	177.79.18.57	homeaway	Chutei uma qualquer com 1/3 de probabilidade.	homeaway	Chutei 1/3 com a mesma probabilidade	Não entendi bem o propósito e também um pouco cansativa.
21	11/06/2017	177.192.197.145	airbnb	Airbnb e 9flag tem descrições muito semelhantes. escolho Airbnb pelo profitStructure.	airbnb	A quantidade de dados é muito maior no Airbnb do que nos demais.	Acho interessante o estudo, mas acho que a informação precisa ser melhor apresentada, e se possível os dados comparativos precisam se aproximar um pouco mais.
22	11/06/2017	187.14.238.18	9flats	I don't know...	airbnb	looks like it's the best according to the opinions I read	It helped me to make a better choice based on the opinion of people who used the services.
23	27/06/2017	187.102.145.186	airbnb	It's the only one that I knew beforehand.	airbnb	Even though it's much more commented, it kept higher rates of positive evaluation.	Terrible data presentation and approach of asking. TOO MUCH data spit in my face without really understanding what is being asked.
24	27/06/2017	200.156.27.28	airbnb	Airbnb is well known, but it is indifferent because need to try the platform	airbnb	Most popular	It confirmed what I already knew.
25	28/06/2017	186.205.119.89	homeaway	IT'S MORE EASY	homeaway	MORE COMFORTABLE TO ME.	REALLY AMAZING.
26	29/06/2017	200.165.167.10	homeaway	HomeAway offers a longer service duration and accept more payment methods.	homeaway	HomeAway seems to offer a better service but has a smaller marketing than Airbnb.	Interesting, different of any other that i've participated. Somewhat confuse at first look but actually good.