

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**ABORDAGEM DE SUPORTE À PRIVACIDADE
EM APLICAÇÕES MÓVEIS CONSIDERANDO A
EXPECTATIVA DE USUÁRIOS**

JOSÉ SANTIAGO MOREIRA DE MELLO

São Carlos - SP
Julho/2018

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**ABORDAGEM DE SUPORTE À PRIVACIDADE EM
APLICAÇÕES MÓVEIS CONSIDERANDO A
EXPECTATIVA DE USUÁRIOS**

JOSÉ SANTIAGO MOREIRA DE MELLO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação. Área de concentração: Redes e Sistemas Distribuídos e Redes de Computadores.
Orientador: Prof. Dr. Sergio Donizetti Zorzo

São Carlos - SP
Julho/2018

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

B248e Mello, Josê Santiago Moreira de
Abordagem de suporte ê privacidade em aplicaíões
m+veis considerando a expectativa das pessoas / Josê
Santiago Moreira de Mello. -- São Carlos : UFSCar,
2018.
115 p.

Dissertação (Mestrado) -- Universidade Federal de
São Carlos, 2018.

1. Abordagem de privacidade. 2. Privacidade m+vel.
3. Expectativa de uso. 4. Expectativa de coleta. 5.
Aplicações móveis. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de Dissertação de Mestrado do(a) candidato **JOSE SANTIAGO MOREIRA DE MELLO**, realizada em **04 de julho de 2018**.

Prof^(a). Dr^(a). Sergio Donizetti Zorzo
(UFSCar)

Prof^(a). Dr^(a). Vania Paula de Almeida Neris
(UFSCar)

Prof^(a). Dr^(a). Ronson Eduardo de Grande
(BU-Can)

Certifico que a defesa realizou-se com a participação à distância do membro Robson Eduardo de Grande, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof^(a). Dr^(a). Sergio Donizetti Zorzo
Presidente da Comissão Examinadora
(UFSCar)

Dedico este trabalho primeiramente a Deus, por ser essencial em minha vida, me iluminando nos momentos de angústias e dificuldades, ao meu pai, José Moreira de Mello Sobrinho, e à minha mãe, Laura de Araújo Moreira. Vocês fazem parte desta importante conquista.

AGRADECIMENTOS

Agradeço eternamente a Deus pela minha vida, por ter me concedido saúde, força e disposição para concluir esta importante etapa em minha vida. Sou imensamente grato ao meu orientador, Professor Doutor Sérgio Donizetti Zorzo, por ter confiado e me dado a oportunidade de alcançar este momento tão importante. Agradeço a esta universidade e a todos os professores que direta ou indiretamente proporcionaram esta conquista. Sou grato, também, à Faculdade Fatec Ourinhos, pelo apoio e incentivo. Por fim, ao velhos e novos amigos que por diversas vezes tiveram participações importantes nesta jornada e principalmente à minha esposa, Thalita Cristiane Ramos, pelos constantes estímulos e pela compreensão nos momentos de dificuldades deste trabalho. Tenho certeza de que sem você de nada ou muito pouco teria sido capaz.

Ama-se mais o que se conquista com esforço.

Benjamin Disraeli

RESUMO

Privacidade é uma preocupação importante no contexto das aplicações móveis, uma vez que essas aplicações manipulam uma grande quantidade de informações confidenciais e pessoais dos usuários por meio de seus dispositivos móveis. Pesquisas anteriores apontam que a manipulação de informações por parte das aplicações é muitas vezes desconhecida pelo usuário, ou então, quando conhecida, o motivo dessa coleta não é suficientemente claro. Diversas ferramentas automatizadas foram propostas para analisar o comportamento de aplicações móveis em relação à privacidade. No entanto, essas ferramentas não levam em conta a percepção dos usuários sobre o que eles acreditam que seja coletado e com qual propósito. Nesse sentido, este trabalho apresenta uma abordagem para privacidade, aplicada no contexto móvel, na qual é considerada a expectativa das pessoas sobre os aspectos de coleta e uso de informação pelos aplicativos de *smartphones*. O estudo relata os resultados de uma pesquisa feita com 163 usuários de *smartphones* que foi utilizada como base para entender como as expectativas das pessoas podem impactar seus sentimentos subjetivos (níveis de conforto). É apresentado, também, o serviço *WePrivacy*, composto principalmente de um aplicativo móvel, conceito que implementa a ideia de privacidade e responde à questão de pesquisa discutida nesse trabalho, ou seja, como considerar a expectativa das pessoas na avaliação e suporte à tomada de decisões sobre privacidade no uso de aplicações móveis. Os resultados encontrados sugerem que os usuários não sabem apontar quais informações suas são coletadas e para qual propósito serve a coleta. Os resultados apontam, também, que, quando se conhecem os propósitos da coleta, os usuários tendem a se sentir mais confortáveis.

Palavras-chave: Abordagem de privacidade; Privacidade móvel; Expectativa de uso; Expectativa de coleta; Aplicações móveis.

ABSTRACT

Privacy is the cause of an important concern in the context of mobile applications, once these applications manipulate a countless number of the users' confidential and personal information through their mobile devices. Previous researches have shown that the manipulation of information through the applications is commonly unknown to the user, or, when known, the reason for this collection is not clear enough. Several automated tools have been proposed to analyze the performance of mobile applications regarding their privacy. However, these tools don't take into consideration the users' perception about what they believe to be collected and the purpose of this collection. Concerning this, the current work presents an approach for privacy applied in the mobile context, in which people's expectation of the aspects of collection and the use of information by *smartphone* applications are considered. This study shows the results of a research performed with 163 *smartphone* users which was used as a basic resource to understand how people's expectations may impact their subjective feelings (levels of comfort). *WePrivacy* service is also presented, mainly composed of a mobile application, a concept which implements the idea of privacy and responds the question of research discussed in this work, that is, how to consider people's expectation in the evaluation and support to the decision making on privacy in the use of mobile applications. The results found suggest that the users don't know how to point which of their personal information are collected and to which purpose the collection serves. The results also show that when the purposes of the collection are known, the users tend to feel more comfortable.

Keywords: Privacy approach; Mobile privacy; Use expectation; Collection expectation; Mobile applications.

LISTA DE FIGURAS

Figura 1.1 - Tempo gasto com interação nos dispositivos - (KPCB, 2017).	14
Figura 2.1 - Conjuntos dos tipos de dados.	25
Figura 3.1 - Uso do <i>TaintDroid</i> na avaliação de aplicativo móvel - Adaptado de Enck et al. (2014).	35
Figura 3.2 - Consulta ao serviço de <i>DNS</i> para o domínio <i>piriform.com</i>	35
Figura 3.3 - Arquitetura de funcionamento do <i>PPMark</i> (PONTES; ZORZO, 2016). ...	38
Figura 3.4 - Representação de modelo mental (BRAVO-LILLO et al., 2011).	43
Figura 4.1 - Distribuição dos participantes por cursos de graduação.	49
Figura 4.2 - Tela para seleção dos aplicativos pelos participantes do estudo.	52
Figura 4.3 - Tela do questionário da etapa (III): Expectativa de coleta/ propósito.	54
Figura 4.4 - Questão optativa para escolha do nível de conforto.	55
Figura 4.5 - Tela do questionário de avaliação da expectativa de uso.	56
Figura 5.1 - Modelo mental do estudo com participantes do grupo expectativa/propósito.	66
Figura 5.2 - Modelo mental do estudo com participantes do grupo perspectiva de conforto.	67
Figura 6.1 - Arquitetura do serviço <i>WePrivacy</i>	73
Figura 6.2 - Modelo relacional do banco de dados do serviço <i>WePrivacy</i>	82
Figura 6.3 - Fluxo de transformação para preparação das análises.	84
Figura 6.4 - Telas iniciais e de análise do <i>WePrivacy</i>	87
Figura 6.5 - Telas do fluxo de participação para o grupo expectativa de coleta/uso.	89
Figura 6.6 - Telas do fluxo de participação para o grupo perspectiva de conforto. ...	90
Figura 7.1 - Tela de exibição das análises do <i>WePrivacy</i>	93
Figura 7.2 - Tela de privacidade para a instalação de aplicativos em sistema <i>Android</i> - Adaptado de Felt et al. (2012).	94

LISTA DE TABELAS

Tabela 3.1 - Análise estática para detecção de execução de códigos remotos maliciosos (POEPLAU et al., 2014).....	33
Tabela 4.1 - Valores críticos associados ao grau de confiança da amostra. Adaptado de Cochran (1977).	47
Tabela 4.2 - Distribuição dos participantes do estudo.....	49
Tabela 4.3 - Lista dos aplicativos móveis mais comuns em 2015 – Adaptado de Annie (2015).....	50
Tabela 5.1 - Aplicativos e recursos com menores expectativas.	60
Tabela 5.2 - Comparativo entre expectativa e o comportamento real do aplicativo. .	63
Tabela 5.3 - Comparativo entre os níveis de conforto para os grupos de participantes.	64
Tabela 7.1 - Comparação de resultados pela perspectiva de usabilidade.	97
Tabela 7.2 - Comparação de resultados pela perspectiva de efetividade.	97
Tabela 7.3 - Resumo das análises de usabilidade e efetividade.....	98

LISTA DE QUADROS

Quadro 2.1 - Práticas justas de uso de informações - Adaptado de OECD (2013)...	23
Quadro 3.1 - Itens de monitoração do TaintDroid - Adaptado de Enck et al. (2014).	36
Quadro 3.2 - Avaliação efetuada com o uso do TaintDroid de aplicações móveis para sistema Android (ENCK et al., 2014).	37
Quadro 3.3 - Exemplo de sequências n-gram.	39
Quadro 3.4 - Rótulo de privacidade (PONTES; ZORZO, 2016).	40
Quadro 6.1 - Algoritmo Python para processamento dos documentos de privacidade.	76
Quadro 6.2 - Trecho da função <i>main</i> do algoritmo de radicalização – Adaptado de Porter (2001).	80
Quadro 6.3 - Função <i>step1()</i> do algoritmo de radicalização - Adaptado de Porter (2001).	81
Quadro 6.4 - Descrição dos ícones de representação.	86
Quadro 7.1 - Formulário usado na avaliação de interfaces.	95

LISTA DE ABREVIATURAS E SIGLAS

ADM - *Administração de Empresas*

ADS - *Análise e Desenvolvimento de Sistemas*

AGRO - *Agronegócios*

AMCIS - *Americas Conference on Information Systems*

API - *Application Programming Interface*

CPF - *Cadastro de Pessoa Física*

DNS - *Domain Name System*

EAD - *Ensino à Distância*

FIPP - *Fair Information Practice Principles*

FTC - *Federal Trade Commission*

GPS - *Global Position System*

HTML - *HyperText Markup Language*

IC - *Intervalo de Confiança*

ICCID - *Integrated Circuit Card Identifier*

IMEI - *International Mobile Equipment Identity*

IP - *Internet Protocol*

OECD - *Organização de Cooperação e Desenvolvimento Econômico*

PDI - *Pentaho Data Integrator*

PLN - *Processamento de Linguagem Natural*

REST - *Representational State Transfer*

RG - *Registro Geral*

SI - *Segurança da Informação*

SSL - *Secure Socket Layer*

TCLE - *Termo de Consentimento e Livre Esclarecido*

TF-IDF - *Term Frequency - Inverse Document Frequency*

TI - *Tecnologia da Informação*

URL - *Uniform Resource Locator*

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO.....	13
1.1 Contexto.....	13
1.2 Objetivos do trabalho	16
1.3 Organização do trabalho	18
CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA.....	19
2.1 Privacidade.....	19
2.2 História da privacidade	20
2.3 Tipo de dados.....	24
2.4 Modelos mentais	26
CAPÍTULO 3 - TRABALHOS RELACIONADOS	30
3.1 Considerações iniciais.....	30
3.2 Mecanismo de permissão do <i>Android</i>	31
3.3 Análise das aplicações móveis.....	32
3.4 Análise do modelo mental	41
CAPÍTULO 4 - ESTUDO COM OS USUÁRIOS.....	44
4.1 Considerações iniciais.....	44
4.2 Seleção dos participantes	45
4.3 Perfil dos participantes	48
4.4 Questionário eletrônico.....	50
CAPÍTULO 5 - RESULTADOS E DISCUSSÕES	58
5.1 Considerações iniciais.....	58
5.2 Menores expectativas	59
5.3 Dificuldades na identificação de propósitos.....	62
5.4 Esclarecer o propósito pode atenuar as preocupações	64
5.5 Modelo mental.....	65
CAPÍTULO 6 - PROTÓTIPO DO WEPRIVACY	70
6.1 Considerações iniciais.....	70

6.2	Concepção Lógica.....	71
6.3	Arquitetura.....	72
6.3.1	Comportamento de aplicações.....	73
6.3.2	Armazenamento de dados	81
6.3.3	Processamento de resultados	83
6.4	Interfaces do aplicativo WePrivacy.....	84
	CAPÍTULO 7 - AVALIAÇÃO DO WEPRIVACY.....	92
7.1	Considerações iniciais.....	92
7.2	Condução da avaliação	93
7.3	Resultados da avaliação	96
	CAPÍTULO 8 - CONCLUSÃO E TRABALHOS FUTUROS.....	99
8.1	Considerações iniciais.....	99
8.2	Contribuições de pesquisa	100
8.3	Conclusões.....	101
8.4	Trabalhos futuros	102
8.5	Publicação de resultados	103
	REFERÊNCIAS.....	104
	APÊNDICE A - TELA DE APRESENTAÇÃO DO TCLE.....	110
	APÊNDICE B - MODELO FÍSICO DO BANCO DE DADOS	111

CAPÍTULO 1

INTRODUÇÃO

Este capítulo apresenta o contexto no qual esta pesquisa se insere, expondo a questão de pesquisa, objetivos, justificativas e a forma como o trabalho está estruturado.

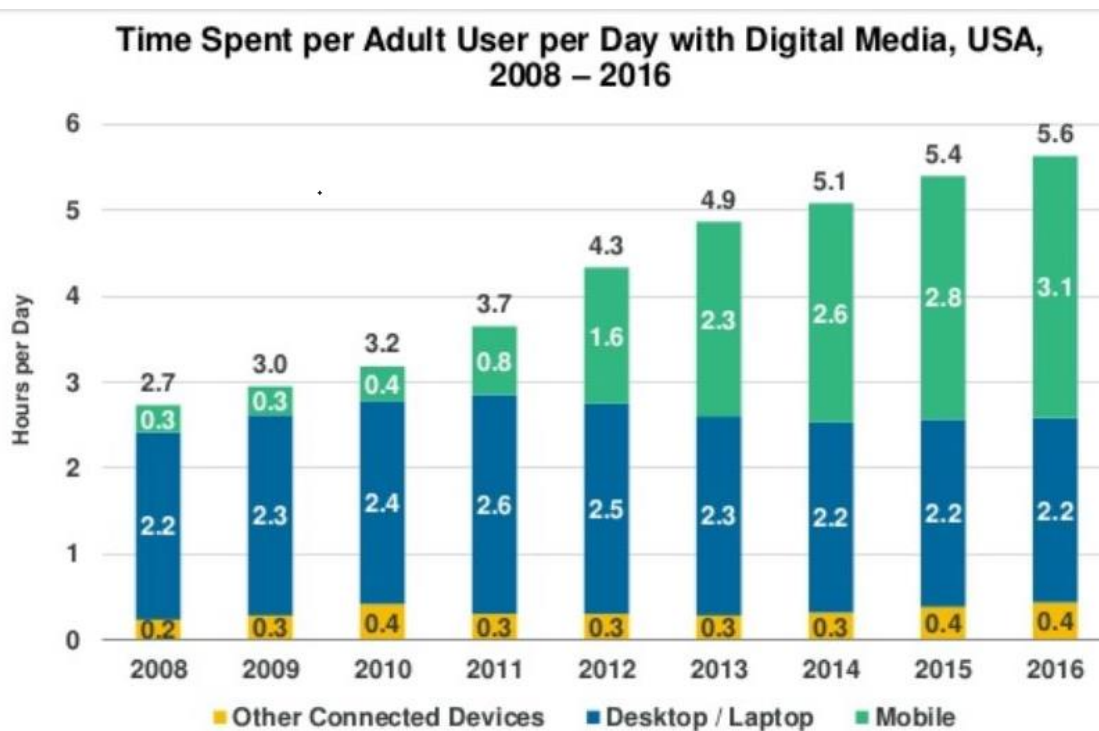
1.1 Contexto

Nos últimos anos, o uso de *smartphones* tem sido cada vez mais presente no cotidiano das pessoas, contribuindo para o crescimento dos aplicativos nesse tipo de plataforma. O uso dessas aplicações proporciona inúmeras vantagens nas atividades do dia a dia para o usuário, como, por exemplo, pagamentos *on-line*, comunicação instantânea, socialização e diversão. Como parâmetro para dimensionar esse crescimento, o relatório anual *Internet Trends*, desenvolvido por Kpcb (2015), aponta um aumento de 58% no número de aplicativos para dispositivos móveis em 2015. A Google Play Store, a loja oficial de aplicativos para a plataforma *Android*, já tinha registrado em 2012 a marca de aproximadamente 700.000 apps entre gratuitos e pagos disponíveis para *downloads* (CARBUNAR; POTHARAJU, 2015).

Outro fator importante a ser analisado nesse contexto está relacionado ao tempo de interação com os sistemas móveis. De acordo com Kpcb (2017), o tempo de interação diária dos usuários com *smartphones* vem crescendo a cada ano, sendo que, a partir de 2014, esse número já superou o dos dispositivos do tipo *desktop* e *notebooks*.

O gráfico apresentado pela figura 1.1 ilustra o tempo gasto diariamente com a interação dos usuários americanos no uso de dispositivos eletrônicos.

Figura 1.1 - Tempo gasto com interação nos dispositivos (KPCB, 2017).



Como observado, o tempo gasto em interações em *desktops* e *notebooks* manteve-se sempre numa linearidade, enquanto o *mobile* vem apresentando crescimento constante ano após ano.

Diante desse cenário, a privacidade com os dados do usuário torna-se uma preocupação, pois muitos aplicativos manipulam (muitas vezes sem o conhecimento dos usuários) diversas informações, tais como localização aproximada, dados de contato, registros de chamada e outras. O controle do acesso a essas informações por parte do usuário é falho ou inexistente, pois não é possível a negociação sobre disponibilizar ou não seus dados para uso do aplicativo, e muitos desses aplicativos não funcionam sem o acesso aos dados do usuário. Ou seja, para poder utilizar os benefícios propostos pelos aplicativos móveis, o usuário tem de concordar em disponibilizar seus dados.

Como consequência, o abuso de permissões no acesso a dados do usuário é muito comum em aplicações de *smartphones*. Alguns apps de jogos, por exemplo, exigem uma conta de cadastro, além de informações como localização, dados pessoais e ID de dispositivos, que podem muitas vezes não ser necessários para a funcionalidade principal do jogo, somente para a formação de um perfil de usuário, muito utilizado para fim de publicidade dirigida. Para a maioria dos usuários a

expectativa principal quanto ao uso de aplicações da categoria jogos é o entretenimento, e muitas vezes não fica claro o motivo de o app necessitar de tantas informações para o cumprimento de sua função principal.

Embora muitos desses aplicativos tenham descrito em sua política de privacidade seus procedimentos quanto à coleta e ao uso de dados do usuário, pesquisas anteriores, como a de McDonald e Cranor (2008), revelam que muitos desses usuários não dão a atenção devida a esses documentos. Essa característica torna as políticas de privacidade um mecanismo ineficiente no propósito de informar ao usuário qual informação o aplicativo obtém e com que propósito.

Outro aspecto relevante desse cenário consiste no fato de o ambiente dos aplicativos móveis ser bem diferente do ambiente dos aplicativos tradicionais *desktops*. O estudo conduzido por Miller, Sumeeth e Singh (2011) revelou que um documento de privacidade escrito para aplicações *desktop* não tem o mesmo efeito em termos de entendimento por parte do usuário de uma aplicação móvel. Isso se dá por vários motivos, entre os quais os mais relevantes são o tamanho da tela e o perfil do uso de um usuário móvel, que comumente está fazendo uso de dispositivo em ambiente mais dinâmico se comparado ao uso das aplicações *desktop*.

De acordo com Nielson e Raluca (2013), o desenvolvimento de qualquer elemento, seja ele uma aplicação ou uma política de privacidade para uma plataforma móvel, deveria ser feito mediante a avaliação das características limitantes do ambiente, como, por exemplo, o tamanho de tela do *smartphone*.

Por essas e outras razões, inclusive culturais, o uso de uma política de privacidade com o objetivo de informar ao usuário como a aplicação móvel irá coletar e manipular seus dados tem se mostrado ineficiente. Muitas vezes, a própria concepção de privacidade que o usuário tem sobre o comportamento de determinada aplicação em relação ao seus dados não está alinhada com o que está descrito na política de privacidade desse aplicativo. Essa divergência está mais relacionada à falta de interesse do usuário pela leitura do documento de privacidade do que ao não entendimento de tal documento (MILLER; SUMEETH; SINGH, 2011).

Em um contexto ideal, se a percepção do usuário quanto ao comportamento de uma aplicação estiver compatível com o documento de privacidade, é certo que as divergências entre expectativa e realidade seriam menores, e com isso haveriam menos problemas de privacidade envolvendo aplicações móveis. No entanto, tal percepção pode ser incompleta ou equívoca. Assim, a existência de um mecanismo

capaz de mostrar ao usuário os principais equívocos entre a expectativa e a realidade sobre o comportamento de um aplicativo contribui para a correção dessa percepção (expectativa x realidade). Esse mecanismo ajuda os usuários na tomada de melhores decisões de confiança sobre o uso (ou não) de aplicações móveis que manipulam suas informações.

É com base nesse contexto que este trabalho apresenta como hipótese a existência de golfos entre o modelo mental feito pelos usuários sobre o uso de suas informações por parte das aplicações móveis e o comportamento real desses aplicativos, cujas regras para coleta e uso das informações do usuário estejam relatadas na política de privacidade. Acredita-se, também, que o desenvolvimento de um mecanismo de avaliação de privacidade, levando em conta a percepção das pessoas, pode contribuir para aproximar o modelo mental que os usuários tenham sobre o comportamento real dos aplicativos móveis.

A questão de pesquisa levantada neste trabalho indaga sobre a possibilidade de se considerar a expectativa dos usuários na avaliação de privacidade para apoiar a tomada de decisão sobre o uso ou não de determinado aplicativo móvel.

A proposta de desenvolvimento de um mecanismo de avaliação de privacidade para aplicações móveis, levando em conta a percepção dos usuários, foi construída por meio de um modelo capaz de extrair do usuário suas expectativas referentes a quais dados são coletados e com que propósito. Essa informação foi comparada com os dados obtidos a partir de uma avaliação do comportamento das aplicações móveis em relação à coleta e ao uso de dados do usuário. A avaliação foi feita com o auxílio de técnicas de análise dinâmicas de aplicações *Android* e com a análise dos documentos de privacidade dos aplicativos avaliados. O propósito dessa etapa foi apresentar as distorções entre a percepção dos usuários sobre privacidade no tratamento de dados e o real comportamento das aplicações móveis.

1.2 Objetivos do trabalho

O objetivo geral desta pesquisa promove uma abordagem de privacidade para aplicações móveis levando em conta a percepção dos usuários. Esta percepção considera aspectos relacionados a qual informação é coletada e o propósito da coleta.

A abordagem tem como contribuição apresentar, por meio de um serviço denominado *WePrivacy*, os principais conflitos entre a percepção do usuário e o comportamento real da aplicação.

O objetivo específico do trabalho é a necessidade da condução de uma pesquisa com o usuário mediante a aplicação de um questionário eletrônico, com o propósito de obter sua percepção referente ao comportamento das aplicações. Essa percepção leva em conta a expectativa do usuário sobre qual informação é coletada e a que fim serve esta coleta. Trata-se de um procedimento de pesquisa que contribuirá para a obtenção do modelo mental dos usuários sobre a manipulação de dados no *smartphone* por parte das aplicações.

O comportamento real da aplicação móvel foi obtido por meio de técnicas de análise dinâmica e Processamento de Linguagem Natural (PLN). A análise dinâmica avalia o comportamento de determinada aplicação enquanto está em execução. Sua condução nesta pesquisa teve como propósito avaliar quais recursos do dispositivo móvel são utilizados pela aplicação. Com o auxílio desse mecanismo, obteve-se quais informações do usuário são coletadas por meio dos recursos dos dispositivos móveis.

A avaliação das políticas de privacidade dos aplicativos analisados é outra importante etapa na busca do objetivo geral desta pesquisa. Sua condução foi feita por meio de técnicas de processamento de texto em linguagem natural, como *Term Frequency-Inverse Document Frequency* - TF-IDF, combinada com o algoritmo *Rabin Karp*. Essa técnica já foi utilizada na pesquisa de Pontes e Zorzo (2016), com o objetivo de apresentar as configurações de privacidade presentes nas políticas de privacidade dos *websites* em um formato tabular, usando a abordagem de rótulo criada por Kelley et al. (2009). Para o contexto desta pesquisa, a avaliação das políticas de privacidade foi utilizada na obtenção dos propósitos de uso de determinada informação coletada.

O resultado da avaliação das políticas de privacidade e da análise dinâmica, em conjunto com a expectativa do usuário, compuseram a base da abordagem de privacidade proposta nesta pesquisa. Os resultados conflitantes, nos quais, na percepção do usuário, o comportamento de uma aplicação foi diferente do mencionado na política de privacidade, foram apresentados em uma interface gráfica de usuário por meio de um novo aplicativo móvel construído como prova de conceito para a abordagem de privacidade apresentada nesta pesquisa.

1.3 Organização do trabalho

Este trabalho está organizado em oito capítulos principais, sendo este Capítulo 1 o responsável por contextualizar o tema, apresentar a questão de pesquisa, os seus objetivos gerais e específicos.

O Capítulo 2 apresenta a fundamentação teórica detalhada sobre as tecnologias e conceitos envolvidos no desenvolvimento deste estudo. O Capítulo 3 discute os principais trabalhos relacionados ao nosso tema.

O Capítulo 4 apresenta o estudo feito junto aos usuários de uma comunidade acadêmica com o objetivo de obter as suas percepções em relação ao comportamento de coleta de dados pessoais manifestado pelos aplicativos móveis selecionados. Este estudo compõe a base de dados inicial do serviço *WePrivacy* proposto.

O Capítulo 5 debate sobre os resultados e apresenta discussões sobre as análises realizadas. O capítulo 6 apresenta o protótipo da aplicação móvel *WePrivacy*, desenvolvida como prova do conceito de privacidade discutido neste trabalho.

O Capítulo 7 discorre sobre a avaliação do protótipo feita junto a uma comunidade acadêmica com o objetivo de validar a proposta apresentada. O Capítulo 8 apresenta as conclusões deste trabalho, bem como as contribuições de pesquisa e as propostas de trabalhos futuros.

Por fim, são apresentadas as referências utilizadas ao longo do trabalho.

CAPÍTULO 2

FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo introduzir conhecimentos específicos que são parte deste trabalho. Para tanto, são detalhados conceitos sobre a privacidade e os fatos que determinaram os diversos aspectos relacionados ao contexto da privacidade, tais como privacidade nos meios de comunicação, privacidade territorial e privacidade das informações. Também são apresentados os tipos de dados e como estão relacionados a este trabalho. Por fim é apresentada a concepção dos modelos mentais na tomada de decisão.

2.1 Privacidade

O conceito de privacidade difere, dependendo do contexto que se avalie e da vivência das pessoas. Não há, por isso, uma definição única sobre o assunto. De acordo com Westin e Solove (2015), não seria possível uma definição unificada para privacidade, pois é assunto que envolve fundamentalmente questões de valores, interesse e poder, que podem ser diferentes entre indivíduos distintos, ficando mais intensas quando consideradas as diferenças culturais.

O conceito de privacidade dado por Rodotà e Moraes (2008) a descreve como a possibilidade de cada indivíduo ter à disposição mecanismos que permitam controlar o uso de informações que lhes dizem respeito. Essa definição condiz com os objetivos e contribuições propostos nesta pesquisa, uma vez que esses mecanismos têm como propósito auxiliar o usuário na tomada de melhores decisões sobre a privacidade de seus dados em relação ao uso das aplicações móveis.

Diversas outras concepções sobre privacidade podem ser encontradas na literatura especializada, como a definição dada por Langheinrich (2001), que afirma que o conceito de privacidade não é apenas a proteção de dados ou o direito de ser “deixado em paz”, mas o direito de escolher e determinar quais atributos de si serão

usados por outros. Corroborando com essa ideia, Westin (1995) define que privacidade é poder se revelar seletivamente para o mundo. Para os propósitos desta pesquisa, portanto, define-se a privacidade como sendo a capacidade de um usuário saber como suas informações serão coletadas e utilizadas, podendo exercer a opção de escolha e o controle sobre seu uso.

2.2 História da privacidade

As discussões sobre privacidade não são preocupação recente. Michael (1994) relata que assuntos sobre privacidade podem ser encontrados desde o ano de 1361, quando um juiz de paz na Inglaterra estabeleceu medidas preventivas quanto aos atos de intromissão e curiosidade nas comunicações alheias, estabelecendo, assim, a primeira noção de privacidade, intitulada como *Privacidade nos meios de comunicação*.

De acordo com Langheinrich (2001), a ideia de *privacidade territorial* surgiu também na Inglaterra no século 18, quando ficou acordado que sobre nenhuma situação o cidadão poderia ter violados os limites de sua propriedade.

No final do século 19, Warren e Brandeis (1890) apresentaram o conceito de privacidade que atualmente é muito difundido como sendo *o direito de estar sozinho*, motivado em grande parte por uma crítica aos repórteres da época, que tiravam fotos das pessoas sem suas permissões.

Na Segunda Guerra Mundial, com o advento do nazismo, os registros a respeito dos cidadãos eram explorados a fim de permitir encontrar com mais facilidade a população judaica das cidades invadidas. Privacidade voltou a ser assunto amplamente discutido na década de 1960, quando os governos perceberam que o tratamento automatizado de dados poderia ser utilizado para catalogar de maneira mais efetiva seus cidadãos. Diante desse cenário, muitos países europeus passaram a criar suas próprias leis de proteção aos dados, buscando prevenir qualquer abuso no uso das informações armazenadas, fato que levou à origem da ideia de *privacidade das informações* (LANGHEINRICH, 2001).

Essas classes de privacidade foram estabelecidas quanto aos seus aspectos legislativos, ficando consolidados principalmente nas leis definidas ao longo do tempo. Tais aspectos são definidos atualmente por muitas nações como *direitos constitucionais*. De acordo com Bhaskar e Ahamed (2007), é a *privacidade das*

informações que dá origem a grande parte dos desafios de ordem legislativa envolvendo a privacidade atualmente.

Posteriormente, na década de 1970, em decorrência dos acontecimentos da década anterior, já relatados, houve grandes esforços na geração de legislações com o propósito de proteção à privacidade. Pode ser destacada como resultado desse esforço a lei americana de privacidade de 1974¹, que apresentou a concepção de práticas justas de uso de informações (*Fair Information Practice Principles* - FIPPs). Essas práticas tiveram um papel relevante na proteção das informações, pois influenciaram de forma significativa as leis e políticas de organizações regulatórias em todo o mundo, como, por exemplo, a comissão federal de comércio americana (*Federal Trade Commission* - FTC)², que adota essas práticas na elaboração das diretrizes que buscam prevenir e eliminar práticas anticompetitivas, como monopólio e cartel.

Em princípio, as FIPPs tiveram sua concepção baseada no trabalho de Westin (1967), que originalmente as definiu em sete categorias, distribuídas da seguinte forma:

1. **Abertura e transparência:** Esta categoria orienta práticas no sentido de não manter registros do usuário de maneira secreta, ou seja, o usuário deve ter ciência de tudo o que for coletado. Há, ainda, orientações sobre a necessidade de fornecer ao usuário meios de aprendizagem sobre o uso de suas informações pessoais, como, por exemplo, uma política de privacidade.
2. **Individualidade:** O assunto desta categoria define práticas relacionadas não somente à possibilidade de conhecer quais dados são coletados. Também orienta para a existência de mecanismos pelos quais o usuário possa contestar a validade e requisitar correções desses dados.
3. **Limite de coleta:** Diz respeito às práticas de coleta de dados dos indivíduos, orientando para coleta somente das informações relevantes e necessárias para alcançar os objetivos propostos. Esta categoria também estabelece que o dado

¹ <https://www.justice.gov/opcl/privacy-act-1974>

² FTC: Agência americana responsável pela proteção e defesa dos direitos e interesses nas relações de consumo (FTC, 2017).

coletado deve ser mantido apenas pelo tempo necessário para cumprir os objetivos especificados. A legislação aplicável deve ser avaliada, pois impacta no tempo de retenção do dado coletado.

4. **Qualidade dos dados:** Esta categoria define a orientação sobre a garantia da qualidade dos dados, ou seja, é necessário prover mecanismos que possam garantir, quando possível, que os dados coletados são precisos, relevantes, oportunos e completos.
5. **Limitação de uso:** O uso de dados coletados deve ser unicamente com as finalidades especificadas no mecanismo de notificação ao usuário. O compartilhamento dos dados não deve ser realizado para outros fins que sejam incompatíveis com a finalidade para a qual o dado foi coletado. Exemplo: É vedado para uma empresa que tenha coletado do cliente seu número de Cadastro de Pessoa Física (CPF) utilizá-lo para realizar uma consulta aos órgãos de proteção ao crédito, não sendo essa a finalidade inicial da coleta. O fato de a empresa dispor dessa informação não lhe dá o direito de utilizá-la da maneira que desejar.
6. **Segurança dos dados:** Esta categoria orienta práticas para proteções dos dados (em todos os tipos de mídias), por meio dos controles de segurança da informação adequados contra riscos como perda, acesso ou uso não autorizado, destruição, alteração ou divulgação não intencional ou inadequada. Esses riscos podem ser mitigados com a adoção de uma política de segurança da informação.
7. **Responsabilidade:** As práticas referentes à responsabilidade dizem respeito a orientações sobre o cumprimento dos princípios propostos, fornecendo treinamento a todos os que manipulam de forma direta os dados dos indivíduos. Orienta também condutas de fiscalização do uso real dos dados coletados, para demonstrar a conformidade com esses princípios e a todos os requisitos de proteção de privacidade aplicáveis.

Internacionalmente, diversas legislações passaram a adotar interpretações das FIPPs, ocasionando, em 1980, um rearranjo dessas práticas por parte da Organização

de Cooperação e Desenvolvimento Econômico (OECD)³. O objetivo dessa reestruturação foi internacionalizar um padrão, de forma a prevenir o crescimento de interpretações distintas sobre as FIPPs, adotadas pelas diferentes legislações, que pudessem impactar o crescimento econômico com a imposição de barreiras comerciais (AGRE; ROTENBERG, 1998).

A abordagem da FIPPs que foi tomada como base na condução deste trabalho é a divulgada pela OECD (2013), definindo oito práticas para as FIPPs que são derivadas da concepção inicial de Westin (1967). O quadro 2.1 descreve um resumo de cada uma dessas práticas.

Quadro 2.1 - Práticas justas de uso de informações - Adaptado de OECD (2013)

Princípios	Descrição
Limite de Coleta	Coletar e manter apenas as informações relevantes para alcançar os objetivos propostos.
Qualidade dos Dados	Garantir que as informações coletadas são precisas, relevantes, oportunas e completas.
Especificação do Propósito	Tornar claro o objetivo da coleta especificando seu(s) propósito(s) de uso da informação coletada.
Limitação de Uso	Utilizar as informações coletadas apenas para os propósitos especificados nos mecanismos de notificação.
Garantia de Segurança	Garantir a segurança contra riscos que possam afetar a confidencialidade, integridade e disponibilidade das informações coletadas.
Transparência	Prover mecanismos de notificações ao usuário sobre coleta, uso, manutenção e disseminação das informações obtidas.
Participação Individual	Busca sempre que possível o consentimento individual para coleta, utilização, divulgação e manutenção das informações obtidas.
Controlador de Dados	Auditar o cumprimento dessas orientações para demonstrar a conformidade com os FIPPs e todos os requisitos de proteção de privacidade aplicáveis.

As categorias em destaque no quadro 2.1, *Especificação do Propósito*, *Transparência* e *Participação Individual*, são de especial interesse para este trabalho, pois são práticas de privacidade nas quais as soluções empregadas direcionam o foco ao usuário. Dado que este trabalho leva em conta a percepção do usuário sobre os aspectos de privacidade em aplicações móveis, acredita-se haver uma maior concordância com essas categorias citadas.

Os demais grupos não destacados no quadro 2.1 são de menor interesse neste trabalho, pois as práticas apresentadas têm como foco maior o lado das aplicações,

³ **OCDE:** Organização de cooperação e desenvolvimento internacional, fundada em 1960 na França e composta por 34 países. Dentre os principais objetivos, destacam-se as propostas para discussões de metas visando o desenvolvimento econômico mundial (OECD, 2016).

que no contexto deste estudo são as aplicações móveis. Essas categorias são relacionadas aos esforços tecnológicos empregados nas aplicações e à recente criação de legislações sobre o tema de privacidade, como o Marco Civil da Internet⁴.

2.3 Tipos de dados

Os tipos de dados são de grande relevância ao se discutir privacidade. Goulart (2015) afirma que, tradicionalmente, se faz uma divisão dos tipos de dados baseando-se na avaliação da sensibilidade. Assim, há os dados pessoais propriamente ditos que trazem consigo elementos que identificam (ou podem identificar) uma pessoa. Por outro lado, há os dados sensíveis que, além de identificar as pessoas, revelam elementos mais profundos de sua personalidade.

Observa-se, a partir dessas definições, que todos os dados sensíveis também são dados pessoais. Contudo, o contrário não é possível de ser afirmado, pois nem sempre um dado pessoal pode ser categorizado com dado sensível (GHANI; SIDEK, 2009).

O uso de conjuntos pode representar essa relação entre dado pessoal e dado sensível. Como observado na figura 2.1, o conjunto dos dados pessoais contém os dados sensíveis.

De acordo com Goulart (2015), a violação de dados sensíveis é muito mais prejudicial para o indivíduo, podendo gerar danos mais intensos à sua personalidade. Outra abordagem envolve a consideração de que o mau uso dos dados sensíveis pode trazer maior possibilidade de discriminação do indivíduo (DONEDA, 2006).

Em complemento a essa divisão tradicional (dado pessoal e dado sensível), Ghani e Sidek (2009) definem, ainda, uma terceira categoria, o dado anônimo. As três categorias de dados apresentadas por Ghani e Sidek (2009) são: I. Dados sensíveis, II. Dados pessoais e III. Dados anônimos.

⁴ Lei n. 12.965: Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil (BRASIL, 2014).

Figura 2.1 - Conjuntos dos tipos de dados



- I. **Dados sensíveis:** Segundo Brasil (2015), é atribuída esta classificação a dados que possibilitam algum tipo de discriminação, como convicção religiosa, políticas e filosóficas, bem como dados que revelem origem racial, ética e questões referentes à saúde.

Colaborando com essa ideia, Goulart (2015) define dados sensíveis como sendo aqueles que, além de representar algum dos aspectos mais íntimos dos indivíduos, também têm um potencial de causar danos mais intensos em situações de mau uso.

- II. **Dados pessoais:** De acordo com Goulart (2015), a lei n. 12.965, que trata do marco regulatório civil da internet, embora mencione no texto o termo *dado pessoal*, não informa qual tipo de informação é considerada como tal. Uma definição para esse conceito pode ser encontrada em Brasil (2015), que caracteriza dados pessoais como sendo qualquer informação relativa a um indivíduo singular identificado ou identificável. Define-se, ainda, que um “indivíduo singular identificado ou identificável” é qualquer pessoa que pode ser identificada, direta ou indiretamente, particularmente por referência a um número de identificação ou por um ou mais elementos específicos de sua identidade física, fisiológica, psíquica, econômica, cultural ou social.

Analisando o conceito anteriormente citado, pode-se estender a ideia de dado pessoal não apenas àquele que o próprio titular possua, como Registro Geral (RG) e CPF, mas também dados que possam identificar um indivíduo com recursos de terceiros, como, por exemplo, dados de endereçamento de rede (IP - *Internet Protocol*).

- III. **Dados anônimos:** São dados relativos a um indivíduo que não possa ser identificado, nem pelo responsável pelo tratamento nem por qualquer outra

pessoa, tendo em conta o conjunto de meios suscetíveis a serem razoavelmente utilizados para identificar o referido indivíduo (BRASIL, 2015).

Dados estatísticos, por exemplo, podem ser enquadrados como dados anônimos, pois, ainda que sejam relativos a um determinado indivíduo (ou grupo de indivíduos), não permitem a identificação de seu titular.

Para o contexto deste trabalho, a categoria de dados pessoais é a mais relevante, pois eles representam a maioria dos dados que é coletada pelas aplicações móveis, como, por exemplo, nome, localização aproximada (extraída a partir do GPS - *Global Position System* do aparelho), ID e dados de agenda do dispositivo. Assim, a pesquisa feita com o usuário para obter sua expectativa quanto à coleta e uso de informações teve como foco estudar o grupo de dados pessoais.

2.4 Modelos mentais

A concepção de modelos mentais é originária da psicologia e o seu conceito foi apresentado por Craik (1967) como sendo um mecanismo do pensamento que as pessoas utilizam para entender e explicar o funcionamento das coisas no mundo real. Para o autor, esse mecanismo tem um papel importante na cognição humana, pois representa abstrações dos processos mentais usados pelo pensamento no reconhecimento, classificação e compreensão das coisas.

Posteriormente, o conceito apresentado por Craik (1967) foi aplicado em vários outros domínios, como a construção de interfaces de usuário, apresentada por Norman (2013), na qual o autor define o uso da psicologia para compreender o modelo mental dos usuários na construção de bons projetos de interação humano-computador.

Mansilha (2008) define modelos mentais como sendo a maneira com que cada indivíduo percebe o que está acontecendo à sua volta, como irá se sentir, pensar e agir sobre a situação apresentada. Senge (2002) amplia esse conceito, descrevendo modelos mentais como sendo uma visão ampla de mundo, adquirida por meio de experiências passadas. Servem como suposições que as pessoas utilizam para tomadas de decisões que orientem os seus comportamentos.

De acordo com Mansilha (2008), os modelos mentais são influenciados por quatro origens distintas, sendo elas: (I) Vivência pessoal, (II) Linguagem, (III) Cultura e (IV) Biológica.

- I. **Vivência pessoal:** O histórico do indivíduo modela suas experiências passadas. Características como raça, sexo, nacionalidade, origem étnica, influências familiares e condição social são exemplos de variáveis que moldam essas vivências e por consequência influenciam na constituição de seus modelos mentais individuais.
- II. **Linguagem:** A linguagem é a segunda origem de influência nos modelos mentais, pois é por meio dela que o ser humano estrutura o seu pensamento. Mansilha (2008) a define como um espaço no qual a realidade apresenta-se de modo inteligível e comunicável. A influência da linguagem na formação do modelo mental ocorre no sentido de que cada indivíduo, ao passar uma informação para outro, o faz destacando os pontos mais relevantes referentes à sua percepção.
- III. **Cultura:** A cultura também pode ser considerada como origem de influência para os modelos mentais. Grande parte dos autores consideram essa influência como modelo mental coletivo, pois dentro de qualquer grupo (família, organizações, empresas ou nações) os modelos mentais são desenvolvidos coletivamente, tendo como base as experiências que são compartilhadas entre os indivíduos desses grupos.
- IV. **Biológica:** Está relacionada às limitações fisiológicas que impedem os humanos de perceber certos fenômenos, por exemplo: os sentidos como audição e visão. De acordo com Mansilha (2008), a impossibilidade de perceber esses fenômenos implica na impossibilidade de agir. Enquanto o cão responde a um apito ultrassônico, uma pessoa nem o ouve. Enquanto o morcego se movimenta sem maiores dificuldades na escuridão, as pessoas se perdem. Por conta dessas limitações fisiológicas, foram inventados equipamentos como sonar e radar para expandir o alcance da percepção dos sentidos e, conseqüentemente, a capacidade de ação frente às coisas ao redor.

Chermack (2003) destaca a característica dinâmica dos modelos mentais, ou seja, com o passar do tempo, mudanças podem alterar a formação desses modelos. Um modelo mental que se adequa bem ao contexto atual pode não valer para um contexto futuro (BARR; STIMPERT; HUFF, 1992).

Devido a essa característica dinâmica, Chermack (2003) afirma que os modelos mentais estão em permanente estado de construção, sendo ajustados e refinados pelas mudanças presentes no ambiente nos quais os indivíduos estão inseridos.

Chapman e Ferfolja (2001) apresentam dois aspectos associados aos modelos mentais. O primeiro está relacionado ao fato de os modelos mentais não serem simplesmente um repositório de aprendizado passado, mas também servirem de base para a interpretação e percepção do que está acontecendo ao redor. O segundo aspecto está relacionado à formação dos modelos mentais como sendo um processo apoiado principalmente nas relações sociais. Dessa forma, o grupo, o contexto cultural no qual o indivíduo se insere, os canais de informação e as experiências vividas constituem agentes influenciadores na formação desses modelos mentais.

Para Lim e Klein (2006), modelos mentais são ativos que moldam nossa forma de agir. Assim sendo, a partir dos estímulos externos, representações culturais do objeto e suas próprias representações mentais, o indivíduo forma de maneira geral sua representação de um objeto, que, conseqüentemente, influenciará suas ações com relação a ele. Os autores afirmam que quanto mais importante for o objeto para o indivíduo, maiores poderão ser os significados associados a ele.

Senge (2002) destaca a individualidade dos modelos mentais, mesmo quando associados a um mesmo objeto. Duas pessoas podem observar o mesmo objeto de diferentes perspectivas e, assim, descrevê-lo de modo distinto. Essa característica faz com que indivíduos distintos possam formar modelos mentais diferentes sobre um objeto em comum.

Avaliar a expectativa do usuário quanto à coleta e uso de informação de seus dispositivos móveis é entender o modelo mental que essas pessoas têm sobre o comportamento do aplicativo móvel. Nesse sentido, comparando essa expectativa com o comportamento real da aplicação, será possível avaliar se o modelo mental que as pessoas têm sobre o funcionamento da aplicação móvel está alinhado com o documento de privacidade desta aplicação. Os resultados dessa avaliação também

poderão contribuir para uma análise sobre o nível de entendimento e interesse dos usuários em ler essas políticas de privacidade.

Para Gardner (2005), é possível a redução do efeito golfo, que se refere a distância entre a percepção e realidade de um fenômeno que é apresentado no modelo mental. Para isso é importante que o indivíduo seja alertado com estímulos comumente externos a respeito da sua percepção incorreta sobre determinado aspectos considerados na formação de seu modelo mental.

CAPÍTULO 3

TRABALHOS RELACIONADOS

Este capítulo apresenta os principais trabalhos identificados na literatura que se correlacionam com o tema proposto nesta pesquisa. A seção 3.1 apresenta as considerações iniciais, relatando como o assunto está organizado. Os trabalhos foram agrupados em três grupos de interesse, sendo a seção 3.2 referente aos trabalhos relacionados aos mecanismos de permissão do Android. A seção 3.3 discute trabalhos sobre o processo de análise de aplicações móveis e suas políticas de privacidade. Por fim, a seção 3.4 apresenta uma pesquisa relacionada ao tema modelos mentais.

3.1 Considerações iniciais

A literatura sobre os aspectos envolvendo privacidade na plataforma móvel é extensa. Podem ser encontradas diversas abordagens, com diferentes objetivos e perspectivas para o problema de avaliar a privacidade nesse cenário. Este capítulo apresenta os trabalhos que têm relação mais próxima ao tema desta pesquisa. Para facilitar a organização e entendimento, optou-se por separá-los em três grupos de interesse, *Mecanismo de permissão do Android*, *Análise das aplicações móveis* e *Análise do modelo mental*.

O primeiro grupo é dedicado às pesquisas que apresentam uma visão relativa ao mecanismo utilizado pelo *Android* para gerenciar as permissões de acesso a recursos do dispositivo do usuário, bem como a aspectos de usabilidade sobre esse mecanismo. O segundo grupo aborda pesquisas que têm como objetivo apresentar técnicas e ferramentas para detectar quando uma determinada informação/ recurso do dispositivo móvel é acessada. O terceiro grupo apresenta uma pesquisa cujo foco é estudar o comportamento das pessoas em relação às suas tomadas de decisões, ilustrando como o modelo mental que as pessoas têm sobre os aspectos avaliados é utilizado para apoiar suas decisões de segurança e privacidade.

3.2 Mecanismo de permissão do *Android*

O mecanismo de permissão do *Android* tem dois propósitos. São eles: (I) limitar o acesso dos aplicativos móveis aos recursos dos dispositivos e (II) ajudar os usuários a tomarem melhores decisões de confiança antes de instalar o aplicativo (KELLEY et al., 2012).

Aplicativos móveis obtidos a partir da loja oficial (*Google Play Store*) somente podem acessar recursos do dispositivo se houver a declaração deste acesso no arquivo de manifesto da aplicação. O arquivo de manifesto (*AndroidManifest.xml*) apresenta informações essenciais sobre o comportamento do aplicativo no sistema operacional *Android*, que necessita conhecer essas informações antes de executar o aplicativo (GOOGLE, 2016).

No processo de instalação de qualquer aplicativo oficial, é apresentada ao usuário a tela de permissão *do Android*, listando os recursos aos quais o aplicativo terá acesso. Essa lista é mapeada a partir do arquivo de manifesto da aplicação. Os usuários podem escolher entre instalar o aplicativo com todas as solicitações de permissão atendidas ou não instalá-lo. Uma vez concedidas, as permissões não poderão ser revogadas, a menos que o aplicativo seja desinstalado.

O estudo conduzido por Kelley et al. (2012) apresentou uma avaliação desse mecanismo de permissão, observando aspectos como a percepção e entendimento dos usuários sobre o que é informado pelo mecanismo.

O estudo foi realizado com o uso de entrevistas semiestruturadas com os usuários *Android* e constatou que as pessoas têm atenção limitada para essas telas de permissão, por não compreender o real impacto que ela implica em relação ao controle de sua privacidade.

Essa falta de entendimento faz com que seja difícil para pessoas que não tenham conhecimentos técnicos formarem uma opinião sobre o risco à privacidade ao instalar o aplicativo em seus dispositivos. Os autores afirmam que as informações sobre as permissões são em grande parte ignoradas pelos usuários, que costumam tomar suas decisões sobre o aplicativo baseados no *ranking* das avaliações dos seus outros usuários.

De forma semelhante Felt et al. (2012) avaliaram a usabilidade do mecanismo de *permissão* do *Android*, com foco no nível de atenção que os usuários tinham com os avisos desse sistema. O experimento foi conduzido tendo como base dois grupos de participantes, um via internet e outro observado diretamente na forma presencial. Em ambos os grupos os resultados foram semelhantes à pesquisa de Kelley et al. (2012).

Felt et al. (2012) concluíram que, para ambos os grupos, os avisos emitidos pelo mecanismo de *permissão* do *Android* não eram efetivos em informar o risco de privacidade envolvido com a instalação do aplicativo. Um dos problemas apontados foi o excesso de avisos, fator responsável pela fadiga nos usuários, que desistiam de ler todas as informações e já realizavam a instalação aceitando as solicitações informadas pelo sistema.

A concepção da abordagem de privacidade utilizada pelo *WePrivacy*, apresentado neste trabalho, contribui com as ideias apresentada por Kelley et al. (2012) e Felt et al. (2012), considerando a expectativa dos usuários na avaliação de privacidade. Isso permite estudar um aspecto não abordado nos estudos anteriores, que é a percepção do usuário sobre o que é coletado e o propósito da coleta, e utilizar isso como uma abordagem de privacidade para apoiar tomadas de decisões de confiança com relação à instalação e ao uso de aplicativos móveis.

3.3 Análise das aplicações móveis

Diversos estudos apresentam técnicas e ferramentas úteis para detectar e notificar a coleta de informações feita pelas aplicações móveis. Essas ferramentas são agrupadas em duas categorias, a saber: (I) análise dinâmica. Avalia a aplicação durante sua execução, registrando todos os acessos aos recursos e informações do dispositivo; (II) análise estática, feita quando a aplicação não está em execução. Nesse método de análise busca-se verificar informações sobre a necessidade do uso de algumas APIs (*Application Programming Interface*), conhecidas por coletar informações dos usuários, como, por exemplo, as APIs de autenticações integradas com redes sociais como Facebook ou Google Plus. A análise estática também busca

avaliar se a aplicação executa códigos externos, como uma chamada a um *webservice*.

Webservices, segundo Rezende (2003), são soluções que permitem a integração de sistemas. São comumente utilizadas na comunicação entre aplicações diferentes. Quando a aplicação não implementa nenhum mecanismo de proteção nas chamadas e execução de códigos externos, isso se constitui numa vulnerabilidade, pois permite que um atacante altere a chamada remota feita pelas aplicações, levando-as a executar um código malicioso (*malware*), que expõe a segurança e a privacidade das informações dos usuários.

Nesse contexto, o trabalho de Poeplau et al. (2014) propôs uma ferramenta para a avaliação estática de aplicações móveis com o objetivo de detectar chamadas de códigos externos que possam ser utilizadas para execução de códigos maliciosos. O experimento conduzido pelos autores avaliou 1632 aplicações móveis escolhidas aleatoriamente entre as aplicações com mais de um milhão de downloads da *Google Play Store*.

Os resultados evidenciam que 9,25% das aplicações apresentavam problemas na chamada de códigos remotos, que permitiriam a execução de um código malicioso em substituição ao código original. Como observado na tabela 3.1, o experimento avaliou o problema da chamada de códigos remotos em cinco categorias: *class loaders*, *package context*, *native code*, *APK instalation* e *runtime.exec*. Cada uma dessas categorias representa diferentes formas com que uma aplicação móvel pode executar um código remoto.

Tabela 3.1 - Análise estática para detecção de execução de códigos remotos maliciosos (POEPLAU et al., 2014)

Category	Application in the Category (relative to the whole set)	Flagged Vulnerable (relative to the whole set)
Class loaders	83 (5.01%)	31 (1.90%)
Package context	13 (0.80%)	13 (0.80%)
Native code	70 (4.29%)	0
APK installation	155 (9.50%)	117 (7.17%)
Runtime.exec	379 (23.22%)	n/a
Total	530 (32.49%)	151 (9.25%)

Na abordagem de análise dinâmica, Enck et al. (2014) conduziram um estudo no qual apresentaram a ferramenta *TaintDroid*, um sistema de análise dinâmica capaz de detectar quando uma determinada aplicação acessa algum recurso monitorado do dispositivo móvel. Para isso, a ferramenta utiliza técnicas de marcações denominadas *taint*, implementadas por meio de modificações no código fonte do sistema *Android*.

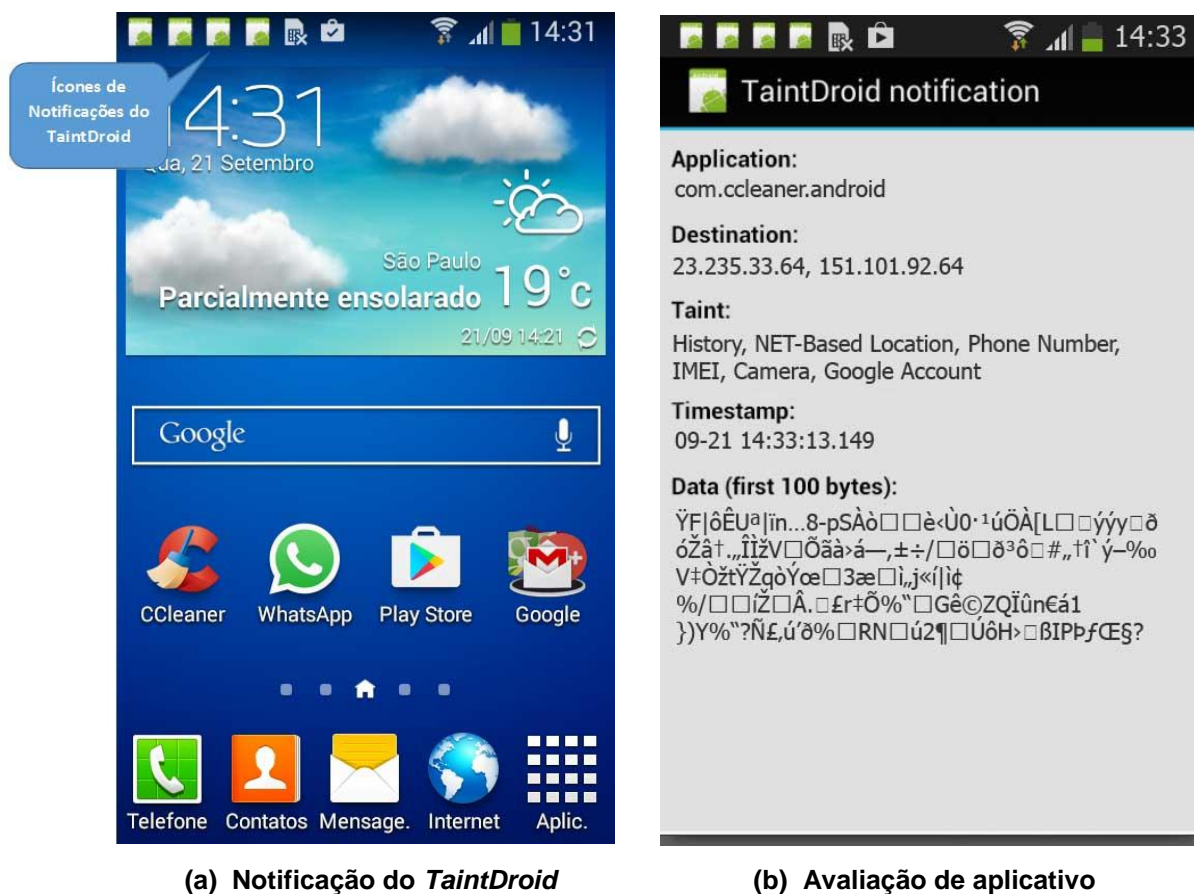
A ferramenta funciona, basicamente, monitorando os recursos do dispositivo durante a execução da aplicação móvel. Para cada acesso aos recursos em monitoração, o *TaintDroid* exibe um ícone de alerta na barra de notificação do *Android*. Ao finalizar a aplicação, essas notificações podem ser consultadas para a obtenção de maiores detalhes, como: quais recursos foram acessados, horário desse acesso e o endereço de rede (*IP*) de destino do envio dessas informações. Também é possível observar uma parcela dos dados coletados/ transmitidos, contudo essa saída exibe a informação sem nenhum tratamento do tipo conversão ou decodificação para torná-la legível.

De acordo com Enck et al. (2014), o tratamento de conversão ou decodificação da seção de informação apresentada pelo *TaintDroid* é inviável. Os principais motivos citados são: os aplicativos móveis podem implementar mecanismos de codificações distintos entre si; o protocolo de comunicação utilizado também pode ser diferente entre as aplicações; diversas aplicações podem implementar mecanismos de criptografia, o que torna essa conversão impraticável.

A figura 3.1a ilustra os ícones de alerta do *TaintDroid* na barra de notificação do sistema. Já a figura 3.1b exibe o detalhamento de um desses ícones, ilustrando o acesso feito pelo aplicativo *Ccleaner*, da empresa Piriform Limited¹.

Como observado na figura 3.1b, o *TaintDroid* informa o acesso aos recursos de *histórico*, *localização baseada em endereço de rede (IP)*, *número do telefone móvel*, *identificação única do aparelho móvel (IMEI)*, *câmera* e dados da *conta google* vinculada ao dispositivo. Também é possível verificar que as informações coletadas são enviadas, por meio da conexão com a internet, a dois endereços de redes (*IP*), sendo eles: 23.235.33.64 e 151.101.92.64.

¹ <https://www.piriform.com/ccleaner>

Figura 3.1 - Uso do *TaintDroid* na avaliação de aplicativo móvel - Adaptado de Enck et al. (2014)

O *Ccleaner* é um aplicativo da categoria utilitário que atua na otimização dos sistemas operacionais Windows, *Android* e Mac Os. Seu funcionamento consiste basicamente na remoção de arquivos temporários desnecessários, como, por exemplo, dados sobre aplicativos já desinstalados do sistema.

Como apresentado pela figura 3.2, uma consulta ao serviço de nomes de domínios (DNS) revela que um dos endereços de rede obtidos com a ferramenta *TaintDroid* pertence ao conjunto de endereços do domínio *piriform.com*.

Figura 3.2 - Consulta ao serviço de DNS para o domínio *piriform.com*

```

C:\Windows\system32\nslookup.exe
Default Server: google-public-dns-a.google.com
Address: 8.8.8.8

> piriform.com

Non-authoritative answer:
Name:     piriform.com
Addresses: 151.101.0.64
          151.101.64.64
          151.101.128.64
          151.101.192.64
>

```

De acordo com Enck et al. (2014), a ferramenta *TaintDroid* tem a capacidade de rastrear, usando as marcações *taint*, de até 30 diferentes tipos de informações no *smartphone* do usuário. Contudo, somente 15 dessas marcações são suportadas atualmente.

O quadro 3.1 apresenta a lista das 15 marcações suportadas atualmente pelo *Taintdroid*. Para o contexto desta pesquisa, as informações de endereço de rede (*taint 0*), lista de contatos (*taint 1*), localização GPS (*taint 4*) e a identificação do dispositivo móvel (*taint 10*) é de especial interesse, pois, segundo estudo conduzido por Hornyack et al. (2011), representam os recursos mais comuns acessados pelas aplicações móveis.

Quadro 3.1 - Itens de monitoração do *TaintDroid* - Adaptado Enck et al. (2014)

Bit de Marcação	Descrição de Marcação
16 - 31	Não Utilizado
15	Dados de histórico
14	Conta Google
13	Serial do dispositivo
12	ICCID (SIM Card ID)
11	IMSI (Identificação Internacional do SIM Card)
10	IMEI (Identificação do aparelho móvel)
9	Dados de SMS enviados recebido
8	Módulo de vibração
7	Módulo de câmera
6	Última localização conhecida
5	Uso de Rede de Dados
4	Localização aproximada por GPS
3	Número do Telefone
2	Microfone
1	Lista de Contatos
0	Localização por ID de Rede (IP)

Com o propósito de avaliar a ferramenta *TaintDroid*, Enck et al. (2014) fizeram um experimento com 30 aplicativos *Android* escolhidos de forma aleatória. Eles concluíram que 66,66% (aproximadamente dois terços) dos aplicativos avaliados manipulam informações dos dispositivos do usuário de maneira imprópria. Em particular, os autores destacaram 15 aplicativos que compartilham a localização do usuário com servidores remotos para fins de publicidade e análises estatísticas para formação de perfil de uso. O quadro 3.2 apresenta os resultados desse experimento.

Quadro 3.2 - Avaliação feita com o uso do *TaintDroid* de aplicações móveis para sistema *Android* (ENCK et al., 2014)

Application [package.name]	Permissions				Info sent		
	Location	Phone state	Camera	Microphone	Location	Phone info	IMEI
Babble Book [com.kalincinsky.babble]	✓						
Cestos Full [com.chickenbrickstudios.cestos_full]	✓						
Manga Browser [com.mangabrowser.main]	✓				.		
Movies and showtimes [com.stylem.movies]	✓						
Solitaire Free [com.mediafill.solitaire]	✓				.		
The Weather Channel [com.weather.Weather]	✓						
3001 Wisdom Quotes Lite [com.xim.wq_lite]	✓				.		
Antivirus Free [com.antivirus]	✓	✓			.	.	.
Astrid [com.timsu.astrid]	✓	✓			.		
BBC News listen & tweet [daaps.media.bbc]	✓	✓			.		
Blackjack [spr.casino]	✓	✓			.		
Bump [com.bumptech.bumpga]	✓	✓					.
Children's ABC Animals (lite) [com.mob4.childrenabc.animals]	✓	✓			.		
Hearts (Free) [com.bytesequencing.hearts_ads]	✓	✓			.		
Horoscope [fr.telemaque.horoscope]	✓	✓			.		.
Mabilo Ringtones [mabilo.ringtones]	✓	✓			.		
The directory for Germany [de.dastelefonbuch.android]	✓	✓					
Traffic Jam Free [com.jiuzhangtech.rushhour]	✓	✓			.		
Wertago for Nightlife [com.wertago]	✓	✓			.		.†
Yellow Pages [com.avantar.jp]	✓	✓					.
Knocking Live Video Beta [com.pointyheadsllc.knockingvideo]	✓	✓	✓				.
Layar [com.layar]	✓	✓	✓				o
Pro Basketball Scores [com.plusmo.probasketballscores]	✓	✓	✓		.		
Slide: Spongebob [com.mob4.slideme.qw.android.spongebob]	✓	✓	✓		.		
The coupons App [thecouponsapp.coupon]	✓	✓	✓			.	.
Trapster [com.trapster.android]	✓	✓	✓				.
Barcode Scanner [com.google.zxing.client.android]			✓				
iXmat Barcode Scanner [com.ixellence.ixmat.android.community]			✓				
Myspace [com.myspace.android]			✓				
Evernote [com.evernote]	✓		✓	✓			

✓ = Potential violation; o = Clearly stated in EULA; † Sent the hash of the value.

Os estudos citados anteriormente representam esforços no sentido de avaliar o comportamento real de uma aplicação. As ferramentas propostas não levam em conta a política de privacidade desses aplicativos ou aspectos associados a esses documentos, como por exemplo, o motivo da coleta.

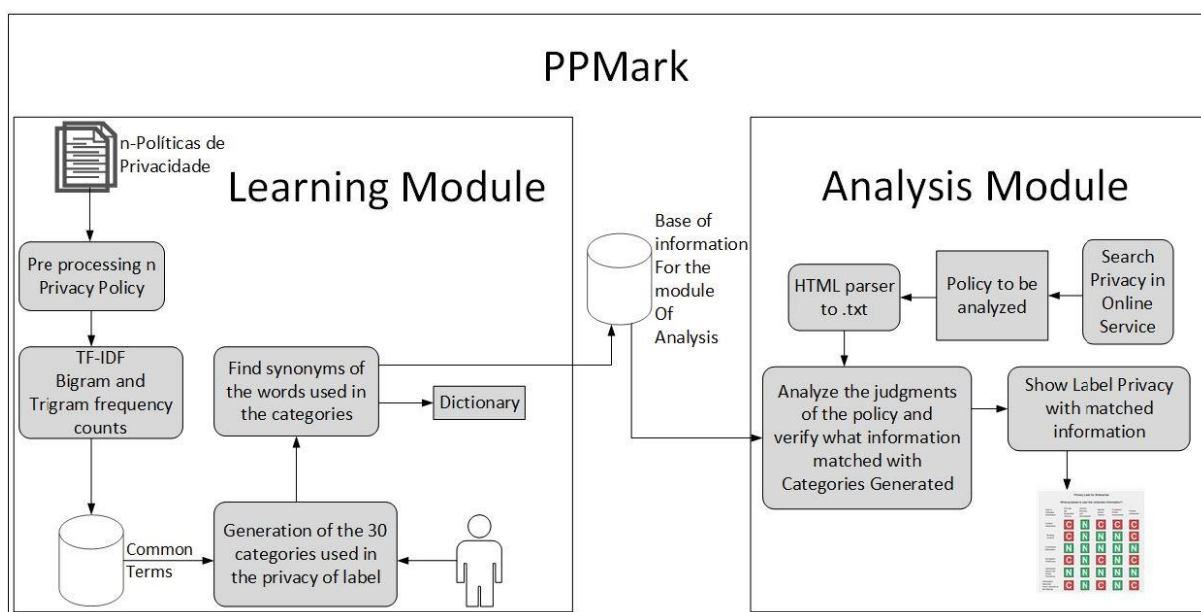
Nesse contexto, Pontes e Zorzo (2016) apresentaram um estudo propondo o uso da ferramenta *PPMark* (*Privacy Policy Mark*), cujo objetivo é apresentar as políticas de privacidade usando uma abordagem tabular denominada rótulo de privacidade. A concepção do rótulo de privacidade foi anteriormente apresentada por Kelley et al. (2009), sendo o *PPMark* uma extensão desse trabalho, no qual os autores apresentam uma proposta para a geração de rótulos de privacidade de maneira automatizada. As técnicas para a geração do rótulo de privacidade são aplicadas nos documentos de privacidade dos websites em que se deseja efetuar a avaliação.

O funcionamento do *PPMark* consiste basicamente em avaliar as políticas de privacidade de websites, por meio do Processamento de Linguagem Natural (PLN),

extraindo informações sobre o que é coletado e o propósito da coleta. Essas informações são utilizadas para a comparação com uma base de conhecimento gerada previamente. O resultado dessa comparação gera as especificações de privacidade que são apresentadas no formato tabular, usando a proposta de rótulo de privacidade apresentada por Kelley et al. (2009).

A arquitetura de funcionamento da ferramenta é dividida em dois módulos, como ilustrado na figura 3.3.

Figura 3.3 - Arquitetura de funcionamento do PPMark (PONTES; ZORZO, 2016)



O módulo de aprendizado é o responsável pela construção da base *n-gram*, utilizada posteriormente pelo módulo de análise da ferramenta.

Um *n-gram* é uma sequência contígua de *n* itens de uma determinada sequência de texto. Esses itens podem representar letras, sílabas, fonemas ou palavras. Os *n-gram* são amplamente utilizados em PLN e mecanismos de reconhecimento de voz.

Um *n-gram* de tamanho 1 é denominado *unigram* (*1-gram*), de tamanho 2 *bigram* (*2-gram*), de tamanho 3 *trigram* (*3-gram*) e assim sucessivamente.

O quadro 3.3 ilustra um exemplo dos *n-gram* de tamanhos *1-gram*, *2-gram* e *3-gram*, extraídos a partir da frase “... ser ou não ser ...”.

Quadro 3.3 - Exemplo de sequências *n-gram*

Texto Base	Unidade	unigram (1-gram)	bigram (2-gram)	trigram (3-gram)
... ser_ou_não_ser ...	Caracter	... S, e, r, _, o, u, _, n, ã, o, _, s, e, r se, er, r_, _o, ou, u_, _n, nã, ão, o_, _s, se, er ser, er_, r_o, _ou, ou_, u_n, _nã, não, ão_, o_s, _ser ...
... ser ou não ser ...	Palavra	... ser, ou, não, ser ser ou, ou não, não ser ser ou não, ou não ser ...

De acordo com Pontes e Zorzo (2016), o *PPMark* utiliza o algoritmo *TF-IDF* para selecionar os *bigram* e *trigram* baseado na consulta dos termos mais comuns presentes nas políticas de privacidade usadas para treinar a ferramenta. Após a obtenção desses termos é feito um agrupamento com o objetivo de criar categorias mais genéricas para representar as informações coletadas e os propósitos das coletas.

O módulo de análise tem como objetivo efetuar a comparação da base *n-gram* com a política de privacidade do website avaliado. Essa comparação é realizada por meio do algoritmo Rabin Karp, que busca a ocorrência de padrões entre os termos do documento de privacidade avaliado e os termos categorizados na base *n-gram*. O resultado dessa comparação é exibido em um formato tabular denominado rótulo de privacidade, no qual é ilustrada a categoria da informação coletada e a que propósito serve a coleta.

O quadro 3.4 apresenta o rótulo de privacidade em formato tabular, que pode ser gerada com o *PPMark*. Os ícones ilustrados pela letra C, destacados pela cor vermelha, representam cenários onde a informação é coletada (grupo da esquerda) para os propósitos especificados (grupo superior). Os ícones ilustrados pela letra N, destacados pela cor verde, indicam cenários nos quais se explicita, na política de privacidade, que a informação apresentada não é coletada e/ou utilizada no propósito destacado.

Como observado no exemplo de relatório do *PPMark* ilustrado pelo quadro 3.4, a coleta de informações pessoais ocorre para prover o serviço solicitado. A leitura de arquivos *cookies* é utilizada para pesquisas e desenvolvimento interno.

Quadro 3.4 - Rótulo de privacidade (PONTES; ZORZO, 2016)

Type of Collected Information	What purpose is used the collected information?				With whom we share the Collect Information?
	Provide the Requested Service	Internal Research and Development	Market Actions	Customer's Profile Assessment	Partner Companies
Contact Information	C	N	N	N	N
Reading Cookies	N	C	N	N	N
Localization Information	N	C	N	N	N
Navigation Preference	N	N	N	N	N
Information About Last Online Purchases	N	C	N	N	N
Information About the User's Activity on the Website	N	N	N	N	N

Os arquivos *cookies* são pequenos documentos de dados enviados de um website para o navegador do usuário. Esses arquivos servem ao propósito de informar ao website a atividade prévia do usuário quando este retornar ao site em um momento posterior.

Além da leitura de arquivos *cookies*, a coleta de informações de localização e informações sobre últimas compras on-line também são obtidas com o propósito de pesquisas e desenvolvimento interno, visando o desenvolvimento de novos produtos ou serviços.

3.4 Análise do modelo mental

O uso de modelos mentais para entender e representar o comportamento humano pode ser observado em pesquisas na área de segurança da informação, como no trabalho de Bravo-Lillo et al. (2011).

Os autores conduziram um estudo para discutir sobre o modelo mental que usuários experientes e inexperientes em tecnologia têm sobre o entendimento dos alertas emitidos pelos aplicativos de diversas categorias.

O objetivo do estudo é explorar os processos psicológicos dos participantes e identificar a linha de raciocínio utilizada para fundamentar as suas ações, bem com os possíveis resultados alcançados pelas decisões escolhidas ao longo do experimento.

A metodologia adotada foi a execução de um experimento com usuários, contendo 30 participantes, sendo 10 deles com conhecimento em computação, denominados como usuários experientes, e 20 com pouco conhecimento relacionado à computação, denominados como usuários inexperientes.

A partir da seleção de um conjunto de mensagens de alerta emitidas por aplicativos de diversas categorias, os participantes eram submetidos a um processo de entrevista semiestruturada, no qual foram questionados sobre o entendimento da mensagem de alerta. Também lhes foi questionado o que aconteceria ao clicar nos botões presentes nas mensagens de alerta e quais ações os participantes recomendariam ser executadas (se aplicáveis) que conduzissem à correção do problema, ou se fariam com que a mensagem não fosse reexibida.

Os autores selecionaram exemplos de 29 mensagens de alerta emitidas por diversos sistemas, dentre eles sistemas operacionais e aplicativos de *software* populares. Essas mensagens foram categorizadas em quatro tipo de avisos: mensagens sobre remoção ou perda de informações; relacionadas à divulgação de informações; execução de códigos maliciosos e confiança em aplicativos ou sistemas de terceiros.

Foi escolhida para participar do experimento, da seleção feita anteriormente, uma mensagem de aviso de cada categoria, sendo: uma mensagem relacionada ao espaço em disco, um alerta relacionado à criptografia de e-mail, uma mensagem relacionada à divulgação de dados do catálogo de endereços do usuário, uma mensagem associada à segurança na abertura de arquivos anexos ao e-mail e uma mensagem relacionada a avisos sobre problemas envolvendo certificados de criptografia de websites (*Secure Socket Layer - SSL*).

As respostas foram obtidas tendo como base a descrição de um cenário fictício, no qual a mensagem de alerta era exibida. Os participantes respondiam questões sobre o entendimento da mensagem e suas ações, baseados em suas experiências passadas na descrição do ambiente, expondo o contexto da mensagem e o texto da mensagem de alerta.

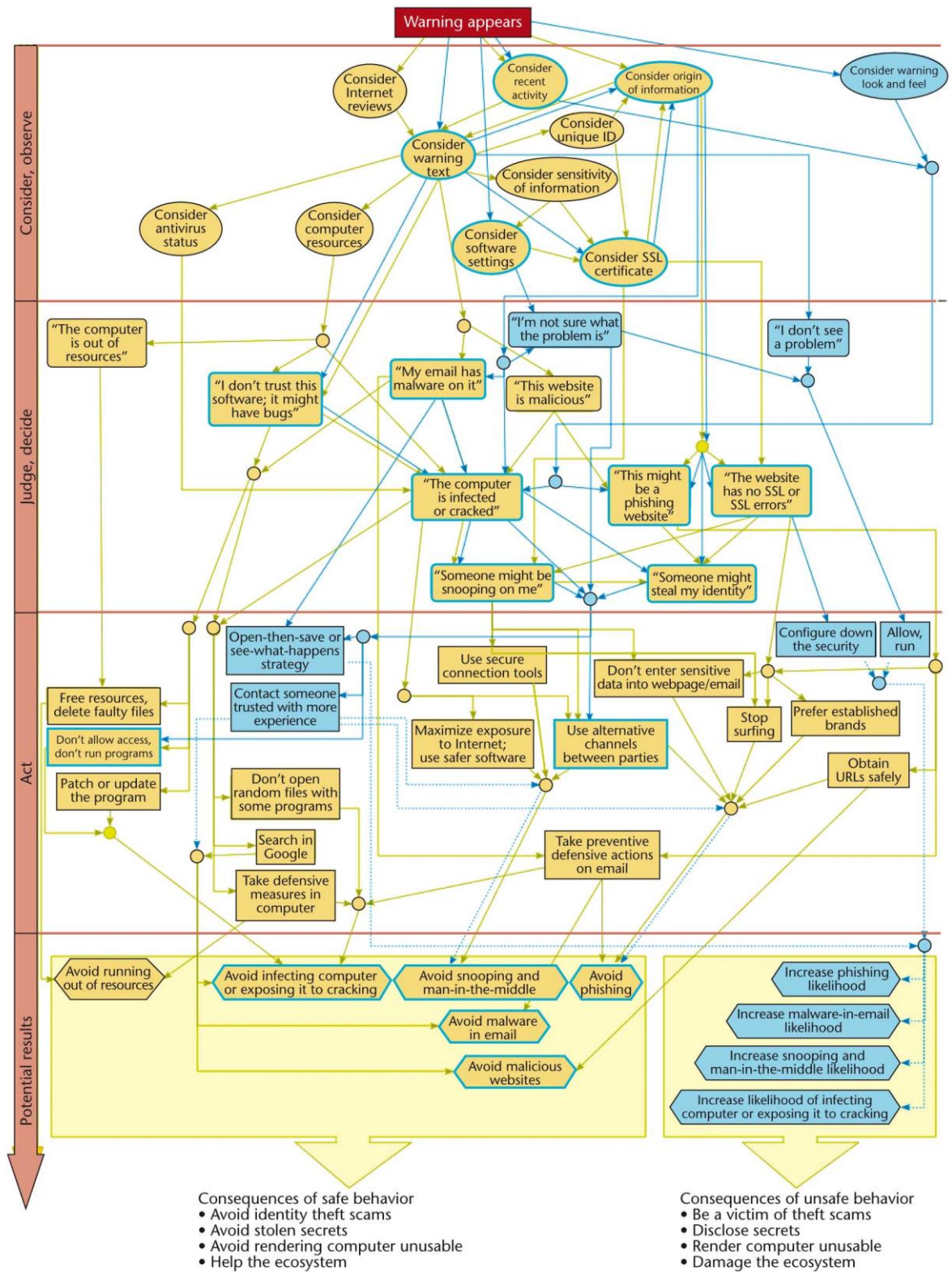
A figura 3.4 representa o modelo mental extraído do experimento. Os fluxos na cor amarela representam as sequências de raciocínio para os participantes experientes (com conhecimento em computação). Os fluxos na cor azul são referentes aos participantes inexperientes (sem conhecimento em computação).

Os elementos circulares representam itens de raciocínio dos usuários utilizados na etapa de observação. Os elementos retangulares de cantos arredondados representam itens de raciocínios de julgamento. Os elementos retangulares sem cantos arredondados representam itens de raciocínios para ações que os participantes recomendariam para solucionar o problema.

Os elementos em forma de losango representam possíveis resultados comportamentais (positivos ou negativos) baseados nos raciocínios apontados no experimento.

O estudo conclui que usuários com diferentes níveis de conhecimento respondem de maneira distinta quando necessitam avaliar e decidir sobre uma mensagem de aviso emitida por sistema computacional. Embora a falta de experiência seja um fator relevante, os autores destacam que mesmo participantes do grupo inexperiente conseguiram formar uma sequência de raciocínio que conduziu a um comportamento potencialmente seguro.

Figura 3.4 - Representação de modelo mental (BRAVO-LILLO et al., 2011)



CAPÍTULO 4

ESTUDO COM OS USUÁRIOS

Este capítulo apresenta o estudo levado a efeito junto aos usuários de uma comunidade acadêmica com o objetivo de obter as suas percepções em relação ao comportamento de coleta de dados pessoais adotado pelos aplicativos móveis selecionados. A seção 4.1 apresenta as considerações iniciais; a seção 4.2 discorre sobre a metodologia estatística usada na seleção dos participantes; a seção 4.3 apresenta o perfil dos participantes selecionados para o estudo e finalmente a seção 4.4 expõe a forma com o sistema eletrônico de coleta foi desenvolvido e apresentado aos participantes.

4.1 Considerações iniciais

Desenvolveu-se, para este trabalho, um estudo junto aos usuários de uma comunidade acadêmica buscando entender as suas percepções em relação ao comportamento de um conjunto de aplicativos móveis pré-selecionados.

Os resultados deste estudo apoiaram a construção de um modelo mental acerca de como estes usuários se sentem em relação ao comportamento de coleta e uso de informações pessoais por parte dos aplicativos móveis.

As informações pessoais tratadas neste contexto referem-se aos dados obtidos pelos aplicativos móveis oriundos dos recursos dos dispositivos, como, por exemplo, agenda de contatos, recurso de GPS e conexões de redes. Tais recursos podem levar à identificação do usuário ou expor a sua localização, como, por exemplo, dados de contatos, identificações de redes conectadas e dados de posicionamentos (GPS).

Os resultados obtidos com este estudo junto aos usuários formam a base de dados inicial do aplicativo mobile *WePrivacy*, desenvolvido como prova do conceito de privacidade discutido neste trabalho.

A privacidade foi abordada, aqui, na forma de expectativa dos usuários sobre o que determinado aplicativo faz e o que ele não faz em relação às informações coletadas, focando principalmente nos aspectos nos quais ele diverge de suas expectativas.

Com esta abordagem, é possível responder à questão de pesquisa deste trabalho, que se refere à possibilidade de considerar a expectativa de privacidade das pessoas para apoiar a tomada de decisão sobre o uso ou não de determinado aplicativo móvel.

Para a elaboração do estudo, foi definido um planejamento de sua execução que envolve desde os métodos estatísticos para a escolha das amostras dos participantes até os procedimentos de coletas de dados. Na seção 4.2, é detalhado o método estatístico adotado para a seleção dos participantes do estudo. A seção 4.3 apresenta os perfis dos participantes que responderam aos questionários do estudo, e por fim, na seção 4.4, é apresentada a forma como o sistema eletrônico do estudo foi construído e apresentado aos participantes.

4.2 Seleção dos participantes

Os participantes do estudo foram selecionados na Faculdade de Tecnologia de Ourinhos (Fatec Ourinhos), dentre alunos dos cursos de graduação tecnológica em Segurança da Informação, Análise e Desenvolvimento de Sistemas, Agronegócios e Gestão Empresarial, sendo este último na modalidade *EAD* (Ensino à Distância).

A população geral considerada para o estudo foi de 968 alunos e todos foram convidados a participar por meio de grupos de e-mails enviados pela própria faculdade e da divulgação feita em sala (para cursos presenciais) por este próprio autor.

A população de um estudo, segundo Morettin e Bussab (2010), é definida como sendo um conjunto finito ou infinito de todos os indivíduos ou objetos que apresentam em comum determinadas características, cujo comportamento interessa analisar. Em nosso contexto, a característica desejada para análise é a expectativa dessas pessoas em relação ao comportamento de um conjunto de aplicativos móveis.

De acordo com Cochran (1977), um contexto ideal seria pesquisar todos os integrantes de uma população. Porém, restrições como tempo, custo e disponibilidade

dos participantes poderiam inviabilizar a condução do estudo com essa abrangência. Devido a tais restrições, normalmente se utiliza uma amostra probabilista estratificada, que, de acordo com Morettin e Bussab (2010), refere-se a uma parcela ou a um subconjunto da população que efetivamente será investigada.

Morettin e Bussab (2010) destacam, ainda, a importância de uma amostra ser escolhida de forma aleatória, tendo cada membro da população igual chance de ser selecionado para compor a amostra do estudo.

Para o cálculo do tamanho de uma amostra (cálculo amostral), Cochran (1977) destaca três variáveis que têm impacto direto sobre seu tamanho, sendo elas: o tamanho da população (quando conhecido), margem de erro e o nível de confiança. O cálculo é definido pela fórmula

$$n = \frac{n_o}{1 + (n_o - 1)/N} \quad (4.1)$$

na qual n representa o número de indivíduos que compõem a amostra e N o número de indivíduos que compõem a população. O valor de n_o é dado pela razão entre o índice de confiança e a margem de erro adotada no estudo. Sua representação é dada pela fórmula

$$n_o = \frac{z^2(k)}{4e^2} \quad (4.2)$$

sendo $z^2(k)$ definido como o valor crítico correspondente ao intervalo de confiança (IC) desejado. A variável e se refere à margem de erro ou à taxa de erro permissível adotada no estudo.

O intervalo de confiança é um indicador de precisão que, segundo Morettin e Bussab (2010), está relacionado a uma medida que estima o quão perto a sua análise, avaliada na amostra, estará do resultado real, quando avaliada na população como um todo.

A margem de erro, também denominada de erro amostral, decorre da diferença entre um resultado obtido na amostra e o verdadeiro resultado obtido na população (MORETTIN; BUSSAB, 2010).

De acordo com Triola (2017), erros amostrais resultam em flutuações amostrais aleatórias, ou seja, duas amostras de uma mesma população podem não obter necessariamente o mesmo resultado.

A taxa de erro ocorre principalmente quando:

- Dados sobre as amostras são coletados, documentados e processados de forma incorreta;
- Utilizam-se instrumentos defeituosos no processo de obtenção das mensurações;
- Há a elaboração de questionários ou formulários com vícios que levam a respostas tendenciosas.

Quanto menor for a taxa de erro que se deseja adotar no estudo, maior será o tamanho da amostra. Essas grandezas (erro amostral e tamanho da amostra) são elementos inversamente proporcionais (TRIOLA, 2017).

Morettin e Bussab (2010) destacam o fato de não ser possível evitar a ocorrência de erro amostral, contudo se pode limitar seu valor por meio da escolha de uma amostra de tamanho adequado.

O processo de cálculo do tamanho da amostra leva em consideração o valor crítico associado ao intervalo de confiança (IC) escolhido. O cálculo deste valor crítico é obtido tendo como base a tabela de probabilidade estatística da distribuição normal Z¹.

Tabela 4.1 - Valores críticos associados ao grau de confiança da amostra - Adaptado de Cochran (1977)

Grau de Confiança	Valor Crítico de Z
90%	1,649
95%	1,96
99%	2,575

¹ http://www.im.ufrj.br/probest/Tabelas_de_probabilidade.pdf

A tabela 4.1, apresentada anteriormente, ilustra os valores críticos de Z para os intervalos de confiança comumente utilizados.

O cálculo amostral adotado neste trabalho foi composto pelos valores de 7% de margem de erro e 95% para índice de confiança.

A equação 4.3 apresenta o detalhamento dos cálculos para a obtenção do tamanho da amostra considerada no estudo.

$$N = 968$$

$$e = 0,07$$

$$z^2(k) = 0,95 \Rightarrow 1,96 \text{ (valor Crítico de Z)}$$

$$n_0 = \frac{z^2(k)}{4e^2} = \frac{1,96^2}{4(0,07)^2} = 196 \quad (4.3)$$

$$n = \frac{n_0}{1 + (n_0 - 1)/N} = \frac{196}{1 + 196/968} = 163$$

Como observado, a amostra calculada para o estudo foi de 163 participantes. Isto representa 16,83% do tamanho total da população considerada, que foi de 968 alunos.

4.3 Perfil dos participantes

Nesta seção é apresentado o perfil dos participantes. A tabela 4.2 apresenta a distribuição de gênero e idade dos participantes que constituíram a amostra do estudo.

Como foi apontado na tabela 4.2, a maioria dos participantes do estudo foi do gênero masculino (64,42%), enquanto as participantes do gênero feminino corresponderam a 35,58%. Esta diferença está relacionada ao perfil de alunos de graduação tecnológica da Faculdade.

De acordo com Sousa et al. (2017), as áreas relacionadas a TI (Tecnologias da Informação), como Ciências da Computação, Engenharia da Computação e Análises

de Sistemas, ainda são predominantemente masculinas, tanto no mercado de trabalho quanto nos cursos de graduação e pós-graduação no Brasil.

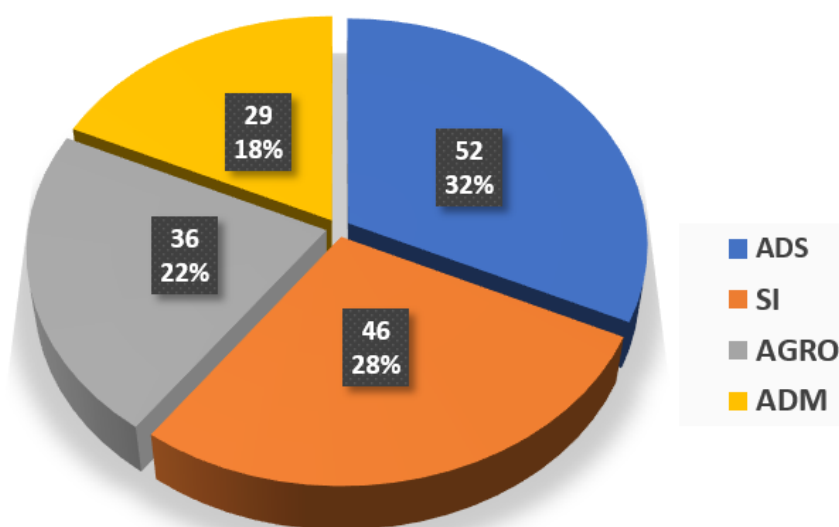
A maior parte dos estudantes que participou do estudo é da faixa etária entre 21 a 25 anos e representa 32,52% do total da amostra.

Tabela 4.2 - Distribuição dos participantes do estudo

Gênero	Participantes	Porcentagem (%)
Masculino	105	64.42
Feminino	58	35.58
Idade		
17 até 20 anos	43	26.38
21 até 25 anos	53	32.52
26 até 30 anos	40	24.54
Acima de 30 anos	27	16.56

Em relação ao curso de graduação dos participantes, a maioria proveio do curso de Análise e Desenvolvimento de Sistemas (52 participantes), representando 32% do total da amostra. A figura 4.1 apresenta o gráfico da distribuição dos participantes em relação aos cursos Análise e Desenvolvimento de Sistemas (ADS), Segurança da Informação (SI), Agronegócios (AGRO) e Administração de Empresas (ADM).

Figura 4.1 - Distribuição dos participantes por cursos de graduação



Administração de Empresas foi o curso com o menor número de participantes do estudo (18%). Acreditamos que tal fato se deve à modalidade do curso ser ensino à distância (EAD), sendo que os participantes foram estimulados a contribuir apenas por meio do e-mail convite. Nos demais cursos, todos na modalidade presencial, além de ter sido enviado o e-mail convite, o autor deste estudo esteve presente em cada sala, intensificando o convite a participação dos estudantes.

4.4 Questionário eletrônico

Com o propósito de obter as expectativas dos participantes sobre as informações coletadas e o motivo da coleta, aplicou-se um questionário eletrônico, contendo um conjunto de perguntas sobre a manipulação de informações dos usuários por parte de um conjunto de aplicações móveis.

Para viabilizar o estudo, foram selecionadas dez aplicações móveis que representam os aplicativos mais comuns (aplicações com o maior número de downloads) em 2015, segundo o relatório da consultoria *App Annie* (ANNIE, 2015).

A tabela 4.3 apresenta a lista dos dez aplicativos *mobile* avaliados em nosso estudo com os participantes.

Tabela 4.3 - Lista dos aplicativos móveis mais comuns em 2015 - Adaptado de Annie (2015)

Ranking	Aplicativo	Companhia
1	WhatsApp Messenger	Facebook
2	Facebook Messenger	Facebook
3	Facebook	Facebook
4	Instagram	Facebook
5	Clean Master	Cheetah Mobile
6	360 Mobile Security	Qihoo 360
7	Skype	Microsoft
8	YouTube	Google
9	UC Browser	Alibaba Group
10	Snapchat	Snapchat

Todos os aplicativos listados na tabela 4.3 foram analisados em nosso estudo pelas suas respectivas versões mais recentes, que estavam disponíveis na loja oficial na primeira semana de novembro de 2016, época em que esta pesquisa foi feita.

O questionário eletrônico usado no estudo foi composto de quatro seções, a saber: (I) *Introdução*, (II) *Pré-seção*, (III) *Expectativa de coleta/ propósito* e (IV) *Perspectiva de conforto*.

Para a condução do estudo, um sistema *on-line* foi desenvolvido para permitir gerenciar de forma eficiente todas as quatro etapas do questionário. O acesso ao sistema *on-line* foi disponibilizado aos participantes por meio do endereço URL (*Uniform Resource Locator*) <http://privacidade.info>, que ficou acessível aos participantes nas três últimas semanas do mês de novembro de 2016.

Inicialmente, ao acessar o sistema *on-line*, os participantes foram apresentados à tela contendo o Termo de Consentimento e Livre Esclarecido (TCLE), no qual era necessário o aceite para iniciar a pesquisa. De forma facultativa, os participantes poderiam receber uma cópia do TCLE em algum endereço de e-mail, se informado. O sistema não armazenava o endereço de e-mail informado, pois isto possibilitaria a identificação do participante.

No Apêndice A deste trabalho pode ser observada uma cópia da tela do sistema eletrônico, apresentando o TCLE utilizado no estudo.

Após o aceite do TCLE, todos os participantes foram redirecionados à (I) *Introdução*, na qual responderam às questões de categorizações com gênero, faixa etária e nome do curso no qual o participante está matriculado.

As questões da seção (I) *Introdução* tinham como objetivo agrupar os participantes para que pudéssemos conhecer melhor o perfil dos pesquisados. Nenhuma das questões usadas no estudo possibilitava a identificação do participante.

Após as questões de categorização apresentadas na seção (I) *Introdução*, os participantes foram direcionados para a etapa (II) *Pré-seção*. Nesta etapa do estudo os participantes foram apresentados a uma lista de dez aplicativos móveis mais populares de 2015, segundo Annie (2015).

Os participantes assinalavam nessa lista, de maneira aleatória, com relação a quais aplicativos eles desejavam opinar sobre o comportamento quanto à coleta e ao uso de dados.

Os estudantes foram informados no e-mail convite e presencialmente pelo autor do estudo, em visitas às salas de aula de cursos presenciais, que a escolha de








muitos aplicativos poderia ocasionar um desconforto em relação ao tempo dedicado a responder a todas as perguntas, pois o tempo estimado foi de aproximadamente dois minutos para cada aplicativo selecionado.

Diante desse cenário, os participantes foram orientados a evitar selecionar mais de três aplicativos, caso não dispusessem de tempo hábil para responder a todas as questões propostas.

A figura 4.2 apresenta a tela desenvolvida no sistema *on-line* referente a esta etapa (*II – Pré-seção*). Os participantes selecionavam os aplicativos sobre os quais desejassem opinar, assinalando uma das três opções em relação à interação com o aplicativo, sendo elas: “*Utilizo este aplicativo*”, “*Já utilizei no passado*” e “*Nunca utilizei, porém conheço o aplicativo*”.

Figura 4.2 - Tela para a seleção dos aplicativos pelos participantes do estudo.

Das opções abaixo, assinale a alternativa que é válida para você em relação ao uso dos aplicativos listados.

	Utilizo este aplicativo.	Já utilizei no passado.	Nunca utilizei porém conheço o aplicativo.
 WhatsApp Messenger	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Facebook	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Facebook Messenger	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Instagram	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Clear Master	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 360 Mobile Security	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Skype	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 YouTube	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 UC Browser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Snapchat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Todos os participantes, inclusive os que assinalaram a interação “*Nunca utilizei, porém conheço o aplicativo*”, tiveram suas opiniões avaliadas no estudo.

De acordo com o estudo conduzido por Bravo-Lillo et al. (2011), e discutido na seção 3.4 do capítulo anterior, mesmo participantes sem experiência com o objeto pesquisado podem fornecer contribuições valiosas para o estudo. Em nosso contexto, avaliamos a interação “*Nunca utilizei, porém conheço o aplicativo*”, fundamentada por essa contribuição dada pelo estudo de Bravo-Lillo et al. (2011).

Após a seleção dos aplicativos, os participantes foram redirecionados a uma das duas etapas seguintes: (III) *Expectativa de coleta/ propósito* ou (IV) *Perspectiva de conforto*. O redirecionamento foi feito pelo sistema *on-line* de forma aleatória entre os participantes, de forma a construir dois grupos de respostas.

Os participantes que responderam às questões referente a um grupo, por exemplo, (III) *Expectativa de coleta/ propósito*, não responderam sobre o outro grupo, (IV) *Perspectiva de Conforto* e vice-versa.

A etapa (III) *Expectativa de coleta/ propósito* constituiu-se de um conjunto podendo conter até três perguntas, nas quais os participantes apontavam a sua expectativa em relação ao que lhes era apresentado referente ao comportamento de cada um dos aplicativos selecionados na etapa (II) *Pré-seção*.

A figura 4.3 apresenta a tela do questionário eletrônico da etapa (III) *Expectativa de coleta/ propósito*, na qual os aplicativos selecionados na etapa anterior foram o Whatsapp Messenger, Facebook, Facebook Messenger e YouTube. Para cada aplicativo selecionado, um conjunto de imagens ilustrativas e um texto descritivo do aplicativo móvel foram apresentados aos participantes. Este conjunto de imagens e o texto descritivo foram extraídos da loja de aplicativos *Google Play Store*.

A etapa (III) *Expectativa de coleta/ propósito* apresentou inicialmente apenas as imagens ilustrativas, o texto descritivo e a questão 1, conforme observado na figura 4.3a.

A questão 1 teve por objetivo obter dos participantes suas expectativas em relação ao comportamento de acesso que alguns aplicativos móveis podem ter em relação ao uso de alguns recursos de seus dispositivos móveis.

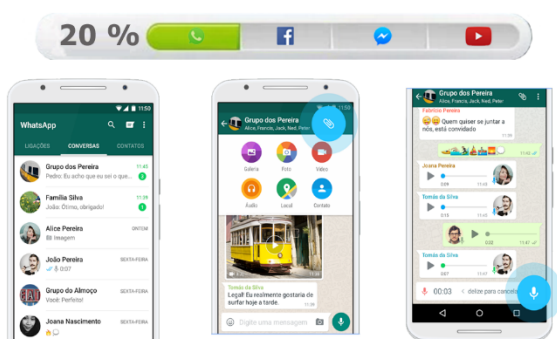
Se o participante responde *Sim*, confirmando acreditar que o aplicativo analisado efetua a coleta da informação apresentada, um conjunto de subquestões é exibido para que o participante possa informar o motivo que ele acredita que o

aplicativo tenha para necessitar do acesso à informação. A figura 4.3b ilustra esse conjunto de subquestões exibido ao participante.

Caso o participante responda *Não*, nenhuma subquestão é exibida. Ao clicar em *enviar*, as respostas fornecidas para a questão 1 e suas subquestões, se houver, serão salvas e o participante será direcionado à segunda e última questão do experimento.

Figura 4.3 - Tela do questionário da etapa (III) Expectativa de coleta/ propósito

Observe atentamente as imagens abaixo e a descrição do aplicativo para responder as questões solicitadas.



Aplicativo: WhatsApp

O WhatsApp Messenger é um aplicativo gratuito para a troca de mensagens disponível para Android e outras plataformas. O WhatsApp utiliza a sua conexão com a internet (4G/3G/2G/EDGE ou Wi-Fi, conforme disponível) para enviar mensagens e fazer chamadas para seus amigos e familiares.

Mude de SMS para WhatsApp para enviar e receber mensagens, chamadas, fotos, vídeos e Mensagens de Voz.

1. O seu aparelho celular armazena diversas informações sobre você. Em relação a isso, considere o aplicativo WhatsApp e assinale abaixo, caso acredite que o aplicativo realiza o uso ou coleta das informações listadas a seguir:

	Sim	Não
Lista de Contatos:	<input type="radio"/>	<input type="radio"/>
Endereço de Rede (Número IP):	<input type="radio"/>	<input type="radio"/>
Localização Aproximada (GPS):	<input type="radio"/>	<input type="radio"/>
Identificação do Aparelho Celular:	<input type="radio"/>	<input type="radio"/>

ENVIAR

(a) Questão inicial

Na segunda questão ilustrada pela figura 4.4, os participantes foram convidados a informar o quão confortáveis se sentiam em permitir que o aplicativo analisado acessasse e coletasse as informações pontuadas na questão anterior.

	Sim	Não
Lista de Contatos:	<input type="radio"/>	<input checked="" type="radio"/>
Endereço de Rede (Número IP):	<input type="radio"/>	<input checked="" type="radio"/>
Localização Aproximada (GPS):	<input checked="" type="radio"/>	<input type="radio"/>
Qual motivo você acredita que o WhatsApp tenha para necessitar dessa informação?		
<input type="checkbox"/> Eu acredito que essa informação seja utilizada para ações de marketing e propaganda.		
<input type="checkbox"/> Eu acredito que essa informação seja necessária para o funcionamento dos principais recursos do WhatsApp.		
<input type="checkbox"/> Eu acredito que essa informação seja utilizada pela equipe do WhatsApp para aprimorar o funcionamento dos recursos da ferramenta.		
<input type="checkbox"/> Eu acredito que o WhatsApp utilize essa informação para compartilhá-la com empresas parceiras com quais o WhatsApp mantém contratos comerciais.		
<input type="checkbox"/> Eu não tenho ideia do motivo pelo qual o WhatsApp necessite dessa informação.		
SALVAR		
Identificação do Aparelho Celular:	<input type="radio"/>	<input type="radio"/>

(b) Subquestão associada à opção “SIM” da questão inicial

Utilizamos uma escala de Likert de quatro pontos, associando os valores negativos para índices que se referem a sentimentos não desejados, como *desconfortável* ou *muito desconfortável*. Associamos valores positivos para índices que expressem sentimentos desejados, como *confortável* ou *muito confortável*.

Diante desse contexto, os valores adotados foram: (-2) *muito desconfortável*, (-1) *desconfortável*, (1) *confortável* e (2) *muito confortável*. Utilizamos estes valores associados às respostas dos participantes, para permitir calcular um índice médio dos sentimentos dos participantes.

Figura 4.4 - Questão optativa para a escolha do nível de conforto.

2. Tendo como base o(s) motivo(s) que você informou na questão anterior, assinale o quão confortável você se sente em permitir que o WhatsApp utilize os recursos listados:

	Muito Confortável	Confortável	Desconfortável	Muito Desconfortável
Lista de Contatos:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Endereço de Rede (Número IP):	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Localização Aproximada (GPS):	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identificação do Aparelho Celular:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

ENVIAR

Optou-se pela não adoção do valor neutro (0) na escala de Likert para evitar o problema de tendência central e por consequência a aversão dos participantes em assinalar os valores extremos, como discutido por Akins (2002 apud ALEXANDRE et al., 2003).

Em relação à etapa (IV) *Perspectiva de conforto*, os participantes selecionados responderam a uma única questão sobre o nível de conforto em relação ao que lhes foi apresentado. A condução dessa etapa foi feita com os participantes selecionados de forma aleatória, após a seleção dos aplicativos feita na etapa (II) *Pré-seção*, conforme apresentado.

A questão apresentada na etapa (IV) *Perspectiva de conforto* teve como objetivo obter o nível dos sentimentos subjetivos, que neste contexto se refere ao nível de conforto que os participantes sentiram quando foram informados sobre o real comportamento de coleta e uso de informações por parte dos aplicativos.

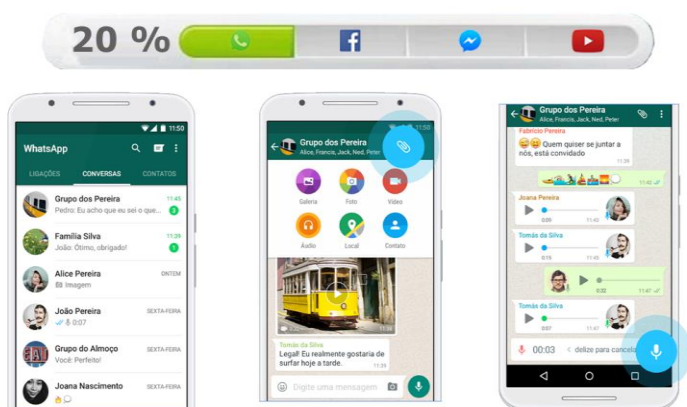
Nessa etapa, os participantes foram avisados de que determinada informação é acessada pelo aplicativo móvel por razões específicas, por exemplo: as informações

sobre *Localização GPS* do usuário sendo utilizadas para fins de *publicidade direcionada*, como verificado no caso do aplicativo *Facebook*.

A figura 4.5 apresenta a questão aplicada aos participantes direcionados à etapa (IV) *Perspectiva de Conforto*.

Figura 4.5 - Tela do questionário de avaliação da expectativa de uso.

Observe atentamente as imagens abaixo e a descrição do aplicativo para responder as questões solicitadas.



Aplicativo: WhatsApp

O WhatsApp Messenger é um aplicativo gratuito para a troca de mensagens disponível para Android e outras plataformas. O WhatsApp utiliza a sua conexão com a internet (4G/3G/2G/EDGE ou Wi-Fi, conforme disponível) para enviar mensagens e fazer chamadas para seus amigos e familiares.

Mude de SMS para WhatsApp para enviar e receber mensagens, chamadas, fotos, vídeos e Mensagens de Voz.

1. Realizamos uma avaliação do aplicativo WhatsApp e de sua política de privacidade. Baseado nisso, descobrimos que esse aplicativo realiza o acesso e coleta de informações de seu aparelho celular.

Algumas dessas informações são:

- **Lista de Contatos** para possibilitar o funcionamento dos principais recursos do aplicativo.
- **Endereço de Rede (Número IP)** para criar um perfil de uso com o objetivo de desenvolver novas funcionalidades.
- **Identificação do Aparelho Celular** para possibilitar o funcionamento dos principais recursos do aplicativo.

Tendo como base essas informações, asinale abaixo o quão confortável você se sente sabendo o que o aplicativo WhatsApp coleta de seu celular e o propósito dessa coleta.

	Muito Confortável	Confortável	Desconfortável	Muito Desconfortável
Lista de Contatos:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Endereço de Rede (Número IP):	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identificação do Aparelho Celular:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

SALVAR

A análise desse grupo de participantes possibilitou avaliar se fornecer informações mais refinadas poderia de alguma forma impactar o nível de conforto das pessoas em relação ao comportamento do aplicativo.

Como foi observado anteriormente, nessa etapa os participantes também foram apresentados a um conjunto de imagens e textos descritivos do aplicativo móvel. Contudo, a questão a ser respondida apresenta a descrição do comportamento do aplicativo no que se refere à coleta e uso de informações do dispositivo móvel.

Adotou-se a mesma escala de Likert usada na etapa (III) *Expectativa de coleta/propósito*, com os mesmos valores numéricos associados, como discutido anteriormente.

Quanto à forma de abordagem desse experimento, caracteriza-se como uma pesquisa quantitativa. Segundo Gil (2010), pesquisas quantitativas consideram que todas as informações no contexto do experimento podem ser mensuráveis para classificação e análises por meio de técnicas estatísticas.

Neste trabalho, além da preocupação com relação à fundamentação estatística para a seleção dos participantes, utilizaram-se técnicas estatísticas como percentagem, médias aritméticas e desvio padrão para analisar e apresentar os resultados obtidos com o experimento.

CAPÍTULO 5

RESULTADOS E DISCUSSÕES

Este capítulo traz os resultados e as discussões sobre o estudo apresentado no capítulo anterior. A seção 5.1 apresenta as considerações iniciais; a seção 5.2 discorre sobre uma análise considerando as menores expectativas dos participantes; a seção 5.3 apresenta uma discussão sobre as dificuldades dos participantes em identificar o motivo que os aplicativos têm para coletar determinada informação; a seção 5.4 discute sobre o alívio das preocupações de privacidade quando o motivo de coleta é suficientemente claro. E, por fim, na seção 5.5 é apresentado o modelo mental da pesquisa desenvolvida com os grupos de participantes.

5.1 Considerações iniciais

Conduziu-se um estudo com uma comunidade específica, conforme a metodologia apresentada no Capítulo 4, com o propósito de avaliar as expectativas de um grupo de pessoas em relação ao comportamento das aplicações móveis. O estudo desenvolvido com os usuários teve como principal característica o entendimento das suas percepções em relação ao comportamento de um conjunto de aplicativos móveis pré-selecionados.

As informações pessoais tratadas neste contexto referem-se aos dados obtidos pelos aplicativos móveis de recursos dos dispositivos móveis que podem levar à identificação do usuário, como, por exemplo, dados de sua agenda de contatos, identificações de redes e posicionamentos do GPS.

Os resultados apoiaram a construção de uma representação para montagem do modelo mental, de forma a ilustrar os fluxos de respostas dos participantes, relacionando o recurso acessado com as percepções subjetivas (níveis de conforto). Isso possibilitou a análise holística da pesquisa feita com os usuários e permite

comparações com futuros trabalhos que venham a abordar o mesmo problema, envolvendo comunidades de usuários distintas.

As respostas dos participantes do estudo compuseram uma base de dados inicial do serviço *WePrivacy*, que é apresentado e discutido no capítulo 6. As análises e discussões apresentadas neste capítulo 5 discorrem sobre os estudos desenvolvidos sobre essa massa inicial de dados e representam o cenário de uma comunidade acadêmica específica.

A generalização dos resultados poderá ser alcançada à medida que o serviço *WePrivacy* for sendo utilizado por outras comunidades de pessoas interessadas. Optou-se entretanto, neste primeiro momento, pela limitação do escopo da comunidade pesquisada, em virtude da fundamentação estatística adotada no estudo, conforme apresentado no capítulo 4.

5.2 Menores expectativas

Em nossa primeira análise, avaliamos as respostas dos participantes referentes ao grupo expectativa de coleta/propósito. Buscou-se identificar quais pares (aplicativo/recursos acessados) tinham a mais baixa expectativa de coleta apresentada pelos participantes.

Para cada par (aplicativo e recurso acessado) fez-se a agregação dos dados e calculou-se a porcentagem de participantes que apresentaram a expectativa correta em relação ao comportamento real do aplicativo.

Esta análise desconsiderou as respostas incorretas em relação às expectativas dos participantes e o comportamento do aplicativo. O objetivo foi identificar no grupo expectativa de coleta/propósito quais aplicativos/recursos representavam as mais baixas expectativas. Isto possibilitou identificar os cenários que mais surpreenderam os participantes.

A avaliação de quais recursos eram acessados pelos aplicativos do estudo foi feita por meio da análise dos registros de *log* da ferramenta *TaintDroid*. A finalidade (propósito) para a qual os dados oriundos desses recursos eram destinados foi verificada por meio de uma análise manual na política de privacidade de cada aplicativo estudado.

Os participantes pontuaram, no estudo, o quão confortáveis se sentiam em relação ao comportamento que eles acreditavam que o aplicativo teria. Essa autoavaliação permitiu analisar qual foi o nível de conforto desse grupo. Avaliou-se o nível de conforto por meio da média aritmética. Os valores associados à escala de Likert adotada para o cálculo foram: + 2 para muito confortável, +1 para confortável, - 1 para desconfortável e -2 para muito desconfortável.

A tabela 5.1 apresenta um resumo dos pares (aplicativos e recurso acessado), no qual pelo menos 20% dos participantes afirmaram que o aplicativo acessaria o recurso apresentado. Por exemplo: apenas 15% dos participantes tiveram a expectativa de que o aplicativo YouTube teria acesso ao recurso de Lista de Contato. No geral, este grupo de participantes sentiu-se entre desconfortável e muito desconfortável sobre o uso desses recursos, tendo como média de conforto -1.32.

Tabela 5.1 - Aplicativos e recursos com menores expectativas

Recursos	Aplicativos	Expectativa (%)	Média de conforto
Lista de Contato	YouTube	15%	-1.32
	Clean Master	15%	-1.18
Localização Aproximada (GPS)	Instagram	10%	-1.10
	Snapchat	5%	-0.95
	Skype	13%	-1.35
	YouTube	7%	-1.12
Endereço de Rede (Número IP)	Snapchat	15%	-1.19
	Clean Master	15%	-1.30
Identificação do Dispositivo	360 Mobile Security	9%	-1.09
	UC Browser	5%	-0.90
	Facebook	20%	-1.45

Da mesma forma, apenas 10% dos participantes tiveram a expectativa correta em relação ao aplicativo Instagram acessar a localização do dispositivo por meio do sistema de GPS e 20% das pessoas tiveram a expectativa correta em relação ao aplicativo Facebook acessar o ID do dispositivo. Para esses casos, a média de conforto foi de -1.10 e -1.45, respectivamente.

De modo geral, quando os participantes foram surpreendidos com um acesso a um recurso do dispositivo móvel, raramente acertavam a questão sobre o motivo de a coleta de dados desse recurso ser necessária.

É importante destacar nesta análise que os participantes (grupo de expectativa de coleta/propósito) não foram informados sobre quais recursos os aplicativos

acessavam e nem sobre o motivo que o aplicativo teria para tal acesso. Nessas condições, observamos uma forte correlação negativa ($r = -0,92$) entre a coluna porcentagem de expectativas e a média de conforto, ambas apresentadas na tabela 5.1.

Em outras palavras, a percepção do usuário sobre o aplicativo necessitar do acesso ao recurso e sua subjetividade (nível de conforto) estão inversamente relacionados, ou seja, quanto maior o número de participantes que tinham expectativas corretas em relação a um aplicativo acessar determinado recurso, menores eram seus níveis de conforto.

O cálculo do coeficiente de correlação adotado nesta análise foi o método de correlação de Pearson (r), que, de acordo com Moore (2010), é dado pela fórmula:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]}} \quad (5.1)$$

Em nosso contexto, n é o número de elementos no conjunto analisado, x representa a variável de porcentagem e y a variável referente à média de conforto.

Morettin e Bussab (2010) afirmam que a correlação é uma medida de associação e mede o grau de relacionamento entre duas variáveis. Serve ao propósito de mensurar a direção e a força da relação linear entre duas variáveis comumente quantitativas.

Os valores de correlação Person (r) variam entre uma escala que vai de +1 até -1. Valores positivos desta correlação indicam correspondência na mesma direção do relacionamento. Já valores negativos indicam correspondências em direção oposta ao relacionamento; representam situação de inversão de proporcionalidade, ou seja, o mesmo fator que movimenta o comportamento de uma variável para uma direção o faz em direção contrária para a outra variável correlacionada, como ocorre na situação encontrada em nosso estudo.

Correlação total (-1 ou 1) indica situações de proporcionalidade perfeita. Significa dizer que, dado o valor de umas das variáveis, com esta característica de correlação pode-se obter o valor da outra. Isso estabelece uma relação de causa-efeito na qual o fator que determina o comportamento de uma variável determina

também, em igual intensidade, o comportamento da outra, positiva ou negativa (MORETTIN; BUSSAB, 2010).

Garson (2011) apresenta uma classificação para valores de correção independentemente do sinal (+ ou -), sendo ela: $r=0,10$ até $0,30$, correlação fraca, $r=0,31$ até $0,70$, correlação moderada, e por fim $r=0,71$ até 1 , correlação alta.

Correlações cujo r seja menor que $0,10$ são muito fracas em relação ao índice de correlação. Nessas situações, o ideal é analisá-las como variáveis sem correlação (GARSON, 2011).

5.3 Dificuldades na identificação de propósitos

As análises feitas com base nas respostas dos participantes do grupo de expectativa mostraram que, no geral, esses usuários têm dificuldade de identificar a razão pela qual um aplicativo acessa determinado recurso.

Com os registros obtidos do *TaintDroid* e as consultas às políticas de privacidade dos aplicativos, categorizamos manualmente cada par (aplicativo/ recurso acessado) em quatro categorias, sendo elas: (1) prover o serviço principal; (2) formação de perfil de uso (rastreadibilidade); (3) publicidade e (4) compartilhamento com empresas parceiras. Muitos recursos listados encaixaram-se em mais de uma categoria, como, por exemplo, o aplicativo *Whatsapp*, que utiliza o recurso de lista de contatos para prover o serviço principal, formar o perfil de uso e compartilhar com empresas parceiras.

Comparamos as respostas de nossos participantes do grupo de expectativa em relação aos valores de nossa análise, conforme apresentado pela tabela 5.2. Na maioria dos casos, os participantes não puderam informar corretamente os motivos pelos quais um aplicativo acessou um determinado recurso.

Outra observação encontrada se refere a quando os recursos foram destinados para o fim de prover o serviço principal (funcionalidade *core*). Em geral, os participantes tiveram maiores acertos. No entanto, essa precisão nunca ultrapassou 65%.

Constatamos, também, que o recurso ID do dispositivo foi o que teve mais respostas assertivas quando comparado com os outros recursos. Quando os recursos

eram utilizados para múltiplas finalidades, por exemplo, lista de contato sendo usada para formação de perfil de uso (rastreadabilidade) e compartilhamento com empresas parceiras, as expectativas corretas tendiam a ser muito menores.

A tabela 5.2 apresenta a análise comparativa entre a expectativa dos participantes e o comportamento real que apontamos em nosso estudo.

Tabela 5.2 - Comparativo entre a expectativa e o comportamento real do aplicativo

Recursos	Propósito(s)	Número de Aplicativos	% Respostas Corretas	% Desconheço o Motivo
Lista de Contato	[1]	7	65%	9%
	[3]	6	45%	38%
	[2] + [4]	5	19%	28%
Localização Aproximada (GPS)	[3]	7	44%	15%
	[1] + [2]	3	23%	32%
	[4]	8	27%	19%
Endereço de Rede (Número IP)	[1]+[3]	6	16%	22%
	[2]	3	13%	37%
	[4]	8	38%	46%
Identificação do Dispositivo	[3]	6	41%	26%
	[4]	7	22%	33%
	[1]	10	62%	11%
	[2]	9	20%	27%

A primeira coluna mostra o tipo de recurso acessado. A segunda exibe o resultado apurado com o uso do *TaintDroid* e com a análise da política de privacidade.

Para facilitar a visualização na tabela do propósito para o acesso ao recurso, optamos por criar códigos para representar esses propósitos a que os recursos dos dispositivos são acessados. Os códigos criados são: [1] prover o serviço principal; [2] formação do perfil de uso; [3] publicidade e [4] compartilhamento com empresas parceiras.

A terceira coluna mostra o somatório do número de aplicativos que se encaixaram em cada categoria. Por exemplo: seis aplicativos acessam o endereço de rede (número IP) com o objetivo de prover o serviço principal (1) e a publicidade (3).

A quarta coluna exibe o percentual dos participantes que afirmaram corretamente o propósito de o recurso ser acessado. Por fim, a última coluna mostra a porcentagem de participantes que afirmaram não saber o porquê de o recurso ser acessado.

5.4 Esclarecer o propósito pode atenuar as preocupações

Devido à falta de clareza do porquê do uso dos recursos de seus dispositivos móveis, os usuários têm de lidar com significativos graus de incerteza na tomada de decisões de confiança quanto a instalar e usar um determinado aplicativo móvel.

Em nosso estudo, queríamos observar se oferecendo informações peculiares, especificamente o motivo de o aplicativo acessar os recursos dos dispositivos, haveria alguma influência nos sentimentos dos usuários com relação à privacidade. Para possibilitar essa observação, como já dito, foram criados dois grupos de participantes: o grupo 1 relacionado à condição de expectativa/ propósito e o grupo 2, relacionado à condição de perspectiva de conforto.

Compararam-se as médias de conforto de ambos os grupos pesquisados para cada par (aplicativo/recurso acessado). Observou-se que os quatro tipos de recursos estudados, ou seja, lista de contatos, localização aproximada (GPS), localização de rede e identificação do dispositivo, tiveram variações significativas em relação ao nível de conforto dos participantes. No geral eles se sentiram mais confortáveis quando foram informados sobre o objetivo de um aplicativo acessar determinado recurso.

A tabela 5.3 apresenta essa análise. As diferenças entre as classificações de conforto mais significativas ocorrem para os aplicativos que acessam o recurso lista de contatos. Em todos os recursos, a classificação média de conforto dos participantes na condição de propósito foi superior ao grupo da condição de expectativas.

Tabela 5.3 - Comparativo entre os níveis de conforto para os grupos de participantes

Recurso	Média de Conforto Grupo Expectativa	Média de Conforto Grupo Perspectiva
Lista de Contato	-0.88	-0.22
Localização Aproximada (GPS)	-0.37	-0.15
Endereço de Rede (Número IP)	-0.63	-0.08
Identificação do Dispositivo	-0.46	-0.11

A primeira coluna refere-se ao recurso acessado. A segunda coluna, média de conforto grupo expectativa, apresenta o cálculo da média aritmética para todas as respostas dos participantes avaliados nesse grupo. A terceira coluna apresenta essa mesma análise para o grupo perspectiva.

Esses resultados sugerem que informar os usuários sobre os motivos pelos quais os aplicativos acessam os recursos de seus dispositivos móveis não só lhes dá condições para melhores tomadas de decisões de confiança, como também pode atenuar suas preocupações causadas pelas incertezas.

As diferenças entre as médias de conforto para ambos os grupos foram significativas para os quatro tipos de recursos. A classificação do conforto variou de -2 (muito desconfortável) a +2 (muito confortável).

5.5 Modelo mental

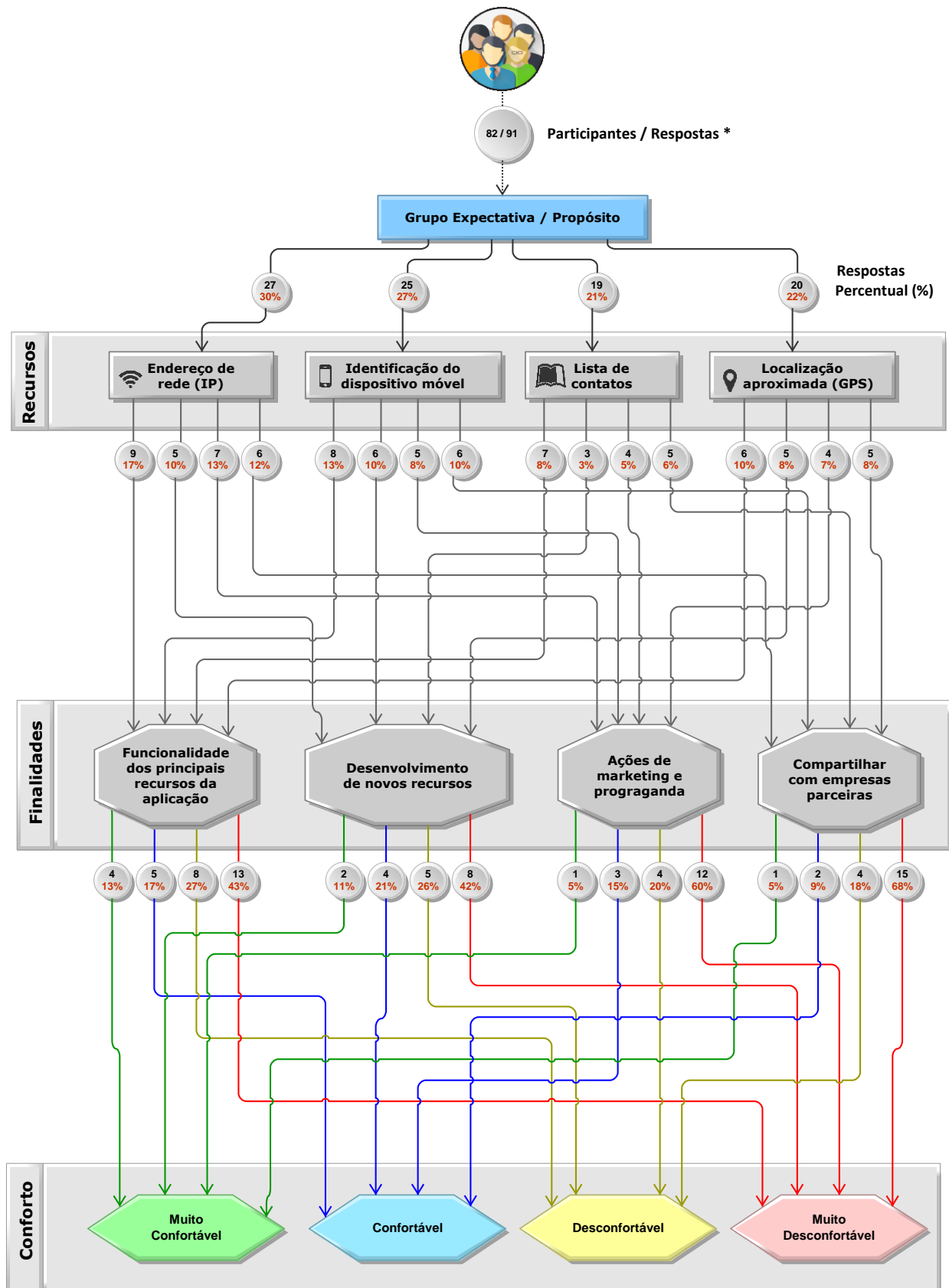
No campo da ciência cognitiva, os modelos mentais são usados para caracterizar as formas pelas quais as pessoas compreendem os sistemas físicos com os quais interagem. Eles servem para explicar o comportamento dos eventos de um sistema, fazer previsões, localizar falhas e atribuir causalidade aos eventos e fenômenos observados (NORMAN, 1987).

Em nosso contexto, construímos uma representação do estudo baseando-nos nas respostas fornecidas pelos participantes, relacionando os eventos de forma a estabelecer uma conexão de causalidade entre os estágios selecionados. As figuras 5.1 e 5.2 apresentam os modelos mentais dos participantes do grupo expectativa/propósito e perspectiva de conforto, respectivamente.

A construção do modelo foi definida pelos fluxos de respostas dos participantes. Estas respostas representam as opiniões para os aplicativos avaliados.

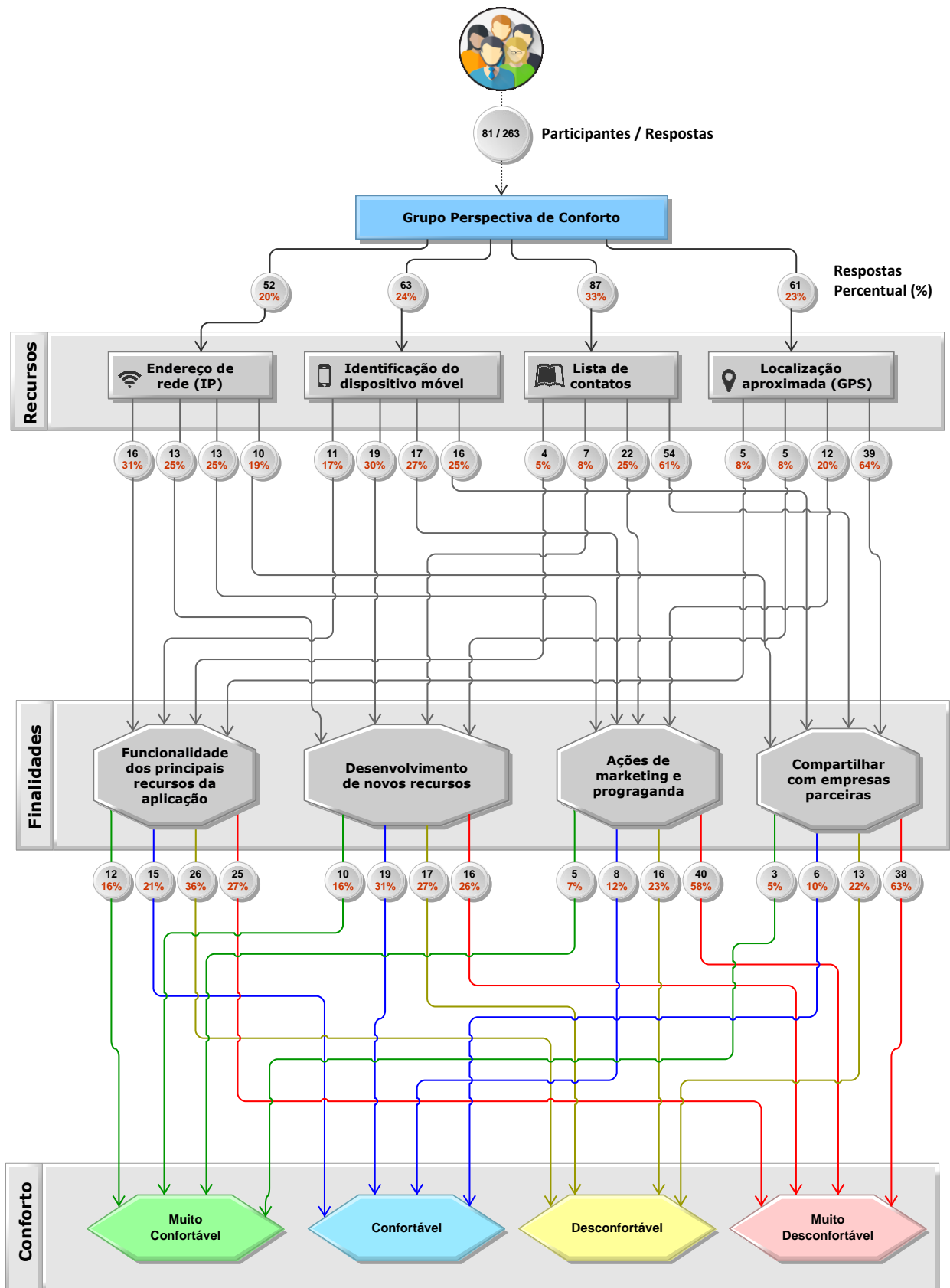
Com o propósito de facilitar o entendimento e a construção da representação do modelo mental, optou-se pela divisão da representação do modelo em três estágios, sendo eles: Recursos, Finalidade e Sentimentos. Os fluxos em forma de seta agregam respostas para cada estágio analisado.

Figura 5.1 - Modelo mental do estudo com participantes do grupo expectativa/ propósito



* Considerando apenas as respostas assertivas em relação ao comportamento avaliado do aplicativo.

Figura 5.2 - Modelo mental do estudo com participantes do grupo perspectiva de conforto



O primeiro estágio, denominado Recurso, apresenta o número de respostas e o percentual de participantes que responderam sobre uma determinada situação. Por exemplo: no grupo expectativa/propósito (Figura 5.1), foram obtidas 27 respostas corretas referentes ao acesso dos aplicativos ao endereço de rede (IP). Isso representou 30% das respostas neste referido grupo. Já no grupo perspectiva de conforto (Figura 5.2), foram obtidas 52 respostas referentes aos aplicativos acessarem o endereço de rede (IP) e isto representou 20% das respostas para esse grupo.

A diferença de respostas entre os grupos está no fato de a abordagem utilizada no modelo para representar as respostas do grupo expectativa/ propósito considerar apenas as respostas corretas.

No estágio 2 analisamos as respostas dos participantes em relação à finalidade (propósito do acesso). No grupo expectativa/propósito, os usuários não foram informados sobre o real propósito do acesso a determinado recurso do dispositivo móvel. Eles responderam com base em suas próprias percepções. Já no grupo propósito, foi informado aos participantes quais recursos eram acessados e com qual propósito. Os participantes, então, apenas sinalizavam quão confortáveis se sentiam com a situação apresentada.

O estágio 3, denominado Conforto, apresenta a maneira como ocorreu a distribuição dos participantes em relação aos seus níveis de conforto. Utilizou-se de cores para representar os fluxos nessas etapas. Fluxos na cor vermelha denotam cenários nos quais os participantes sentiram-se muito desconfortáveis. Por exemplo: no grupo expectativa/propósito (figura 5.1), todas as respostas assertivas em relação ao fato de os recursos serem utilizados para ações de marketing e propaganda, ou seja, 20 respostas ou 60%, revelam que os participantes se sentiram muito desconfortáveis com esse comportamento.

Os fluxos na cor amarela denotam as situações nas quais os participantes sentiram-se desconfortáveis. Já os fluxos na cor azul se referem aos participantes que se sentiram confortáveis. Finalmente, os fluxos na cor verde representam as situações nas quais os participantes sentiram-se muito confortáveis.

O modelo mental para o grupo expectativa/propósito considerou apenas as respostas corretas em relação ao comportamento do aplicativo avaliado. Nesse grupo os participantes não foram informados sobre o comportamento do aplicativo, fato que gerou muitas respostas incorretas. Ao todo, os 82 participantes do grupo forneceram

um total de 253 opiniões (respostas), das quais apenas 91 estavam corretas. Isso representou uma taxa de assertividade de aproximadamente 36%.

A análise dos modelos mentais evidencia uma tendência dos participantes de sentirem-se mais confortáveis quando são informados sobre o comportamento de um aplicativo em comparação ao grupo de expectativa/propósito.

Fundamentamo-nos nessa evidência de que, quando melhor informados sobre os aspectos relacionados à privacidade no uso de aplicações móveis, os usuários se sentem mais confortáveis, propusemos a criação do serviço *WePrivacy*, que tem por objetivo ajudá-los na tomada de decisão de confiança em relação à privacidade.

CAPÍTULO 6

PROTÓTIPO DO *WEPRIVACY*

Este capítulo apresenta o protótipo da aplicação desenvolvida como prova do conceito de privacidade em foco neste trabalho. Tal conceito põe em questão como considerar a expectativa das pessoas para apoiar a tomada de decisão sobre a confiança no uso dos aplicativos móveis. A seção 6.1 apresenta as considerações iniciais e as correlações do assunto com os capítulos já apresentados. A seção 6.2 discute a concepção lógica do protótipo desenvolvido. A Seção 6.3 discorre sobre a arquitetura de serviço proposta para suportar o funcionamento do protótipo e a seção 6.4 apresenta as telas e discute as formas de interações entre o usuário e o protótipo.

6.1 Considerações iniciais

Nos capítulos 4 e 5, identificamos e discutimos que o propósito da coleta de determinada informação dos usuários e a suas expectativas em relação a esse comportamento por parte dos aplicativos são dois fatores chaves que influenciam o dos usuários.

Com base nessa constatação, apresentamos o projeto do *WePrivacy*, um serviço que implementa a ideia de privacidade como expectativa, ou seja, considera a opinião das pessoas para auxiliar novos usuários interessados na tomada de decisão de privacidade em relação ao uso de aplicativos móveis.

O aplicativo protótipo *WePrivacy* foi desenvolvido na plataforma Android e é compatível com vários modelos de *smartphones*. O aplicativo foi construído tendo como base a ideia de ser um serviço de apoio à privacidade. Seu principal objetivo é informar aos usuários interessados a expectativa de outras pessoas em relação ao comportamento de determinado aplicativo móvel.

A base de dados inicial do serviço foi construída a partir do estudo desenvolvido juntamente aos usuários, como apresentamos no capítulo 4 e discutimos no 5.

Duas características direcionaram o projeto do serviço *WePrivacy*. A primeira está relacionada à possibilidade de o *WePrivacy* ser utilizado no contexto real, fora do ambiente acadêmico, permitindo aos usuários tomar melhores decisões de privacidade e confiança no aplicativo no qual estivessem interessados. A segunda característica é a oferta de um serviço que pudesse considerar as opiniões das pessoas sobre os comportamentos de aplicativos móveis, possibilitando-nos, com isso, compormos uma medida de avaliação que apresentasse o nível de conforto dos usuários em relação ao comportamento do aplicativo observado.

Para permitir a expansão a outros domínios de aplicativos e o crescimento da base de dados do ambiente, o *WePrivacy* possibilita aos participantes contribuir com o serviço, respondendo ao questionário fornecido pelo próprio serviço, referente a um app que o usuário tenha instalado em seu dispositivo. Outra forma de contribuir é opinar sobre uma análise apresentada. Neste contexto, a opinião do participante é processada como sendo do grupo de perspectiva de conforto, tendo em vista que o participante observou as análises do comportamento do aplicativo avaliado, apresentadas pelo *WePrivacy*.

A possibilidade de considerar contribuições para prover o serviço é o conceito fundamental do *WePrivacy*. Essa ideia é denominada de *crowdsourcing*, conceito que em português significa contribuição colaborativa ou colaboração coletiva.

A ideia do uso da *crowdsourcing* adotada em nosso protótipo segue a mesma lógica subjacente ao trabalho de Patil, Page e Kobsa (2011), no tocante a incorporar a opinião dos outros na tomada de decisão de privacidade. Contudo, nosso trabalho difere no sentido de considerarmos o uso da *crowdsourcing* ao invés de partir de um círculo social (rede social) para construir a opinião coletiva sobre os aspectos de privacidade.

6.2 Concepção lógica

No geral, os mecanismos de aviso de privacidade, como, por exemplo, a tela de permissão do Android, são projetados para serem analisados pelos usuários antes da instalação do aplicativo no dispositivo. Nesse curto espaço de tempo, os usuários têm informações limitadas e dificuldades para formar o seu modelo mental em relação

aos aspectos de privacidade do aplicativo móvel, uma vez que ainda não tiveram nenhuma interação com ele.

Nesse aspecto, McDonald e Cranor (2008) destacam a falta de interesse do usuário em examinar as políticas de privacidade de um serviço, seja ele um *website* ou um aplicativo móvel, com o objetivo de entender como as questões de privacidade funcionam de fato. Em nosso projeto, utilizamos da concepção lógica apresentada por Patil, Page e Kobsa (2011) no sentido de incorporar as opiniões dos outros para auxiliar na tomada de decisões de privacidade.

6.3 Arquitetura

Tendo em vista a limitação quanto à capacidade de processamento dos dispositivos móveis, foi necessária a adoção de arquitetura que permitisse centralizar o processamento de todas as análises feitas pelo *WePrivacy* em um recurso computacional externo ao dispositivo móvel do usuário.

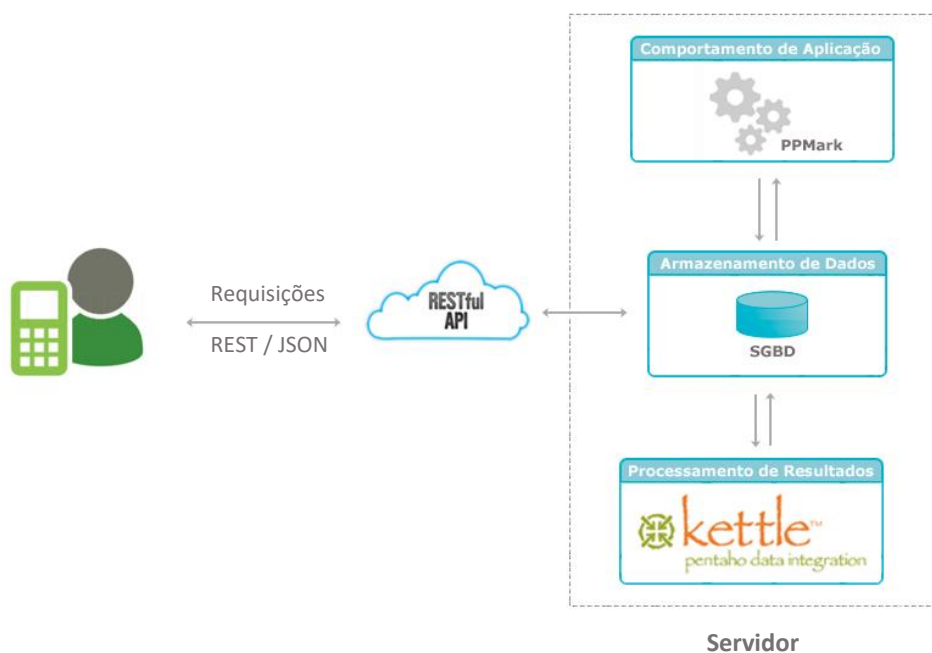
A arquitetura cliente-servidor foi adotada neste contexto, sendo o cliente composto de uma aplicação móvel (*WePrivacy*) que o usuário interessado instala em seu dispositivo. Esta aplicação se comunica com o servidor para fornecer aos usuários as análises e os estudos processados pelo serviço.

Por meio das interfaces de participações, os usuários podem contribuir com o *WePrivacy*, com o app enviando as respostas dos participantes ao servidor para processamento.

A figura 6.1 apresenta os componentes da arquitetura adotada no *WePrivacy*. A comunicação entre o app cliente e o servidor é feita por meio da internet e utiliza-se de um serviço *web* (*web service*) *restfull*, desenvolvido em java para a troca de mensagens entre cliente (aplicativo móvel) e servidor.

Restfull é o protocolo de comunicação baseado na tecnologia REST (*representational state transfer*), comumente utilizada na integração entre sistemas *web* e sistemas *mobile* (LECHETA, 2015).

Figura 6.1 - Arquitetura do serviço WePrivacy



O servidor do *WePrivacy* é o componente da arquitetura responsável pelo processamento dos estudos apresentados aos usuários.

Como ilustrado pela figura 6.1, sua estrutura do lado servidor é composta de três módulos, sendo eles: Comportamento de aplicação, Armazenamento de dados e Processamento de resultados.

6.3.1 Comportamento de aplicações

Este módulo é o responsável pela obtenção da lista de recursos e da finalidade (propósito) do acesso que os aplicativos avaliados pelo *WePrivacy* executam.

A abordagem proposta por Enck et al. (2014) sobre o uso da ferramenta *TaintDroid* apresenta duas limitações que impossibilitam o seu uso neste contexto. A primeira está relacionada à impossibilidade de o *TaintDroid* informar a finalidade (propósito) de determinada coleta. A segunda está relacionada à não possibilidade de automatizar os processos de análise, já que a ferramenta não tem suporte a essa operação.

A automatização do processo de análise do comportamento das aplicações móveis listadas pelo *WePrivacy* é um importante requisito do projeto, sem a qual o módulo demandaria uma intervenção manual para a execução dessa etapa.

Para contornar essas limitações e atender ao requisito de automatização, foram utilizadas técnicas de Processamento de Linguagem Natural (PLN), aplicadas sobre as políticas de privacidade dos aplicativos.

A técnica de PLN adotada nessa etapa foi a implementação proposta por Pontes e Zorzo (2016), denominada *PPMark*.

A proposta do *PPMark* tem por objetivo extrair das políticas de privacidade quais informações são capturadas do usuário e para que elas são utilizadas. Pontes e Zorzo (2016) propuseram a ferramenta para avaliar políticas de privacidade de websites e apresentar os resultados em um formato tabular denominado rótulo de privacidade, como apresentado no capítulo 3, seção 3.3 (análise das aplicações móveis).

Embora o mecanismo tenha sido projetado inicialmente para avaliar a política de privacidade de *websites*, seu uso no contexto deste trabalho foi possível devido aos documentos de privacidade relacionados aos aplicativos móveis serem ou terem versões *on-line*.

O *PPMark* utiliza como mecanismo de análise textual os algoritmos *TF-IDF* e o *Rabin Karp*, sendo o primeiro utilizado para a criação de um grupo de palavras-chaves baseadas na avaliação das frequências dos termos encontrados nas políticas de privacidade e o segundo responsável pela comparação entre as palavras-chaves e o documento de privacidade avaliado. O propósito dessa comparação é estimar uma correlação entre os elementos avaliados que possibilitem identificar qual informação é coletada e o motivo da coleta.

O uso do *PPMark* como mecanismo de PLN apresentou duas limitações para alcançar os objetivos dessa etapa do trabalho. (I) A primeira está relacionada à utilização de especialistas (intervenção humana) para a escolha de quais termos serão relevantes para compor a lista de palavras-chaves que é utilizada posteriormente pelo algoritmo *Rabin Karp*. (II) A segunda limitação está associada ao processo de radicalização (*stemming*), que é o mecanismo responsável pela remoção das variações morfológicas de uma palavra. No *PPMark*, esse processo utiliza a implementação de Orengo e Huyck (2001), que considera apenas textos escritos em

português. Como neste trabalho houve a necessidade de avaliar documentos de privacidade que estavam disponíveis somente em inglês, foi preciso adicionar ao *PPMark* um mecanismo de radicalização capaz de tratar textos escritos também nesse idioma.

I - Limitação ao uso de especialista

A limitação ao uso de especialista na escolha dos termos relevantes para este trabalho tem como objetivo evitar tendências e vícios nos resultados obtidos pelo *PPMark*, uma vez que as palavras-chaves seriam escolhidas com base no conhecimento empírico do especialista. Contudo, o uso do especialista nesta etapa impossibilita a automatização do processo.

Para contornar essa limitação o trabalho adotou como conjunto de palavras-chaves as listadas no resultado de uma avaliação dos termos mais recorrentes em 50 políticas de privacidade dos *websites* mais acessados na categoria computação móvel. Optou-se por esta categoria em virtude da similaridade dos seus termos em relação às políticas de privacidade das aplicações móveis.

A seleção desses *websites* foi feita a partir do site Alexa Internet Inc¹, uma empresa subsidiária da Amazon que oferece serviços relacionados a pesquisas e medições de audiência para *websites* da internet.

Inicialmente, os documentos foram convertidos para o formato texto puro, isto é, removendo as marcações de *HyperText Markup Language (HTML)* presentes nos documentos *on-line*. Para essa finalidade, foi desenvolvida uma aplicação de *Web Scraping*. São classificados como *Web Scraping* mecanismos automatizados capazes de extrair e processar informações de *websites* de forma automática. Eles também são conhecidos como ferramentas de extração de dados e são úteis para automatizar atividades de cópia ou digitação repetitiva relacionada a documentos *on-line*.

Desenvolvida usando a tecnologia *Python*, a ferramenta *Web Scraping* construída copia as páginas de políticas de privacidade e posteriormente as analisa removendo a codificação HTML presente no documento. O quadro 6.1 apresenta o código fonte desenvolvido. As linhas 12 e 13 são as responsáveis por copiar as políticas de privacidade a partir dos endereços de *Uniform Resource Locator (URL)*

¹ http://www.alexa.com/topsites/category;0/Top/Computers/Mobile_Computing

fornecidos no arquivo *lista.url* (linha 5). A linha 18 é a responsável pela conversão em formato texto simples, removendo as codificações HTML do documento eletrônico obtido. Como observado, esse procedimento é feito por meio da biblioteca *BeautifulSoup*², do *Python*.

Quadro 6.1 - Algoritmo *Python* para processamento dos documentos de privacidade

Código Fonte: web_scraping.py

```
1 import urllib
2 from BeautifulSoup import BeautifulSoup
3 from os import mkdir
4
5 url_file = '/root/scripts/lista.url'
6 controle = 0
7
8 arq = open(url_file, 'r')
9 base = arq.readlines()
10 mkdir('pol_priv_text')
11 for url in base:
12     sock = urllib.urlopen(url)
13     raw_html = sock.read()
14     sock.close()
15
16     politica = 'pol_priv_text/politica-' + str(controle) + '.txt'
17     texto = open(politica, 'w')
18     texto.write(BeautifulSoup(raw_html).text)
19     texto.close()
20     controle = controle + 1
21 arq.close()
```

Os documentos de privacidade obtidos nessa etapa foram submetidos ao processo de remoção de termos denominados *stopwords*, que são palavras consideradas irrelevantes na análise de textos, justamente por não expressarem a essência do assunto tratado. Fox (2003) afirma que as *stopwords* normalmente são palavras classificadas como artigos, preposições, pronomes e outras classes de palavras auxiliares. Uma coleção de *stopwords* é denominado como *stoplist*.

De acordo com Fox (2003), os elementos que compõem uma *stoplist* dificilmente são utilizados na avaliação dos textos. Além disso, considerar tais elementos tornaria mais complexo o trabalho dos algoritmos de análise textual.

A *stoplist* utilizada nessa etapa do trabalho foi obtida da proposta de Fox (2003), que apresenta um estudo sobre as palavras mais comuns para *stopwords* em textos em inglês. O resultado do estudo sugere uma *stoplist* composta por 278 palavras que,

² <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

de acordo com o autor, pode ser utilizada para o processamento em documentos escritos em inglês.

Após a remoção das *stopwords*, os documentos foram analisados por um algoritmo de classificação de textos denominado *TF-IDF*, com o propósito de obter o índice de frequência das palavras nos documentos. As palavras foram ordenadas em ordem decrescente pelo índice de frequência. Optou-se por considerar todas as palavras-chaves cujo índice de frequência, calculado com o *TF-IDF*, fosse igual ou superior à média deste índice. Todas as palavras selecionadas compuseram o grupo de palavras-chaves utilizadas pelo *PPMark* em substituição à análise do especialista.

O funcionamento do algoritmo *TF-IDF* consiste, basicamente, na avaliação da importância da palavra (termo) dentro de um conjunto de documentos (*corpus*).

Conforme o número de vezes que um determinado termo aparece no documento, passa a ter maior relevância na análise. Porém, se esse mesmo termo aparece diversas vezes em vários documentos do mesmo conjunto, então esse termo passa a ter menor relevância dentro de todo o conjunto (DUARTE, 2009).

O *TF-IDF* é obtido calculando-se separadamente a frequência do termo (*TF - Term Frequency*) e a frequência do termo invertida no documento (*IDF - Inverse Document Frequency*) e depois, então, multiplicando-se os resultados entre si para atingir o valor final *TF-IDF*.

O *TF* é definido como sendo o número de vezes em que um termo aparece no documento. Contudo, Duarte (2009) afirma que tal frequência deve ser normalizada de forma a evitar tendências nas avaliações de documentos longos. Assim, Duarte (2009) define o cálculo do *TF* como:

$$tf_{ij} = \frac{f_{ij}}{\sum_{i=1}^n f_{kj}} \quad (6.1)$$

na qual f_{ij} é a quantidade de ocorrências do termo t_i no documento d_j e k é a quantidade de termos distintos.

O *IDF* mensura a importância geral do termo no *corpus*. Seu cálculo, segundo Duarte (2009), é o logaritmo do quociente entre o número total de documentos D e o número de documentos que contêm o termo.

Sua fórmula é dada por

$$idf_i = \log \frac{D}{|D_{t_i}|} \quad (6.2)$$

na qual D é o número total de documentos avaliados e $|D_{t_i}|$ representa o número de documentos que contêm o termo t_i .

Assim sendo, o *TF-IDF* pode ser definido como o produto entre *TF* e *IDF*, sendo representado por:

$$TF-IDF = tf_{ij} \times idf_i \quad (6.3)$$

Com o resultado do processamento do algoritmo *TD-IDF*, obteve-se uma lista das palavras mais comuns nos documentos de privacidade avaliados. Essas palavras refletem os termos mais comuns relacionados aos mecanismos e aos dados de coletas descritos nos documentos de privacidade avaliados da categoria de computação móvel.

II - Limitação do processo de radicalização

O processo de radicalização tem um papel importante na descoberta de informações textuais. A principal função desse mecanismo é permitir o mapeamento de palavras que sejam semanticamente e morfologicamente relacionadas, com o propósito de agrupar o maior número de palavras que tenham o mesmo sentido por meio de seu radical. Um exemplo desse processo pode ser analisado observando-se o grupo de palavras: {análise, analisada, analisando, analisado, analisados, analisaram}, no qual, apesar de terem características diferentes, as palavras preservam um mesmo radical, *ANALIS*. A radicalização do termo apresenta a origem primitiva dos vocábulos, isto é, mostra como os termos eram anteriores às variações, como plurais e inflexões verbais (PORTER, 1980).

A proposta do *PPMark* utiliza o processo de radicalização apresentado por Orengo e Huyck (2001), que tem por objetivo efetuar a radicalização de textos escritos em português. Essa abordagem não se aplica, contudo, a alguns documentos de privacidade dos aplicativos que serão avaliados pelo serviço *WePrivacy*, pois seus

documentos de privacidade estão disponíveis apenas em inglês. Nesse contexto, Ebecken, Lopes e Costa (2003) descrevem três métodos de radicalização que podem ser aplicados a textos escritos em inglês: *Método Stemmer S*, *Método Porter* e *Método Lovins*.

Método Stemmer S: Considerado o método mais simples. Seu funcionamento consiste basicamente na eliminação de apenas alguns caracteres ao final das palavras, geralmente sufixos que formam o plural como *ies*, *es* e *s*.

Método Porter: Consiste na localização de diferentes inflexões referentes à mesma palavra e na sua substituição por um radical comum. O algoritmo remove cerca de 60 sufixos diferentes para palavras de língua inglesa e seu funcionamento é baseado nas seguintes fases: redução do plural, troca de sufixos, retirada de sufixos, remoção de sufixos padrões e remoção da vogal “e” ao final da palavra (BASTOS, 2006).

Método Lovins: Proposto por Lovins (1968), é um método de um único passo, que utiliza um algoritmo para remover cerca de 250 sufixos diferentes. Esse método submete cada palavra uma única vez ao processo de remoção de sufixo, retirando o sufixo mais longo conectado a ela. O algoritmo foi criado a partir de um conjunto de exemplos de palavras, usado para formar a lista de regras de Lovins (REZENDE, 2003).

O estudo conduzido por Jivani et al. (2011) comparou as vantagens e desvantagens no uso dos algoritmos citados. O algoritmo de Porter produziu melhores resultados e menor taxa de erro quando comparado com os outros algoritmos de mesma categoria. Sua implementação está disponível em várias linguagens, como Java, C, C#, Perl e Python (PORTER, 2001).

Diante do contexto apresentado por Jivani et al. (2011), o método Porter foi o algoritmo de radicalização utilizado para implementar no *PPMark* o processo de radicalização de palavras em inglês.

Os quadros 6.2 e 6.3 apresentam parcialmente a implantação do algoritmo de Porter usando a linguagem Java.

Quadro 6.2 - Trecho da função *main* do algoritmo de radicalização - Adaptado de Porter (2001)

Código Fonte: porter_stemming.java

```
1 public static void main(String[] args)
2 {
3     char[] w = new char[501];
4     Stemmer s = new Stemmer();
5     for (int i = 0; i < args.length; i++)
6         try
7         {
8             FileInputStream in = new FileInputStream(args[i]);
9             ...
10            ch = in.read();
11            if (!Character.isLetter((char) ch))
12            {
13                ...
14                s.stem();
15                { String u;
16                  u = s.toString();
17                  System.out.print(u);
18                }
19            }
20            ...
21        }
22    ...
23 }
24
25 public void stem()
26 { k = i - 1;
27   if (k > 1) { step1(); step2(); step3(); step4(); step5(); step6(); }
28   i_end = k+1; i = 0;
29 }
```

A chamada do método *main* (linha 1) recebe como parâmetro uma lista de arquivos em que se deseja efetuar o processo de radicalização. A linha 5 executa uma estrutura de repetição (*for*), submetendo cada arquivo ao processamento do algoritmo. As linhas 4 e 14 são as responsáveis, respectivamente, pela criação do objeto tipo *stemmer* e a chamada da função *stem()*, que executa o processo de radicalização em 6 etapas.

Quadro 6.3 - Função *step1()* do algoritmo de radicalização - Adaptado de Porter (2001)

Código Fonte: porter_stemming.java

```

1  private final void step1()
2  { if (b[k] == 's')
3    { if (ends("sses")) k -= 2; else
4      if (ends("ies")) setto("i"); else
5        if (b[k-1] != 's') k--;
6    }
7  if (ends("eed")) { if (m() > 0) k--; } else
8  if ((ends("ed") || ends("ing")) && vowelinstem())
9  { k = j;
10     if (ends("at")) setto("ate"); else
11     if (ends("bl")) setto("ble"); else
12     if (ends("iz")) setto("ize"); else
13     if (doublec(k))
14     { k--;
15       { int ch = b[k];
16         if (ch == 'l' || ch == 's' || ch == 'z') k++;
17       }
18     }
19     else if (m() == 1 && cvc(k)) setto("e");
20 }
21 }

```

O quadro 6.3 apresenta o detalhamento da primeira etapa do processo de radicalização, que ilustra a chamada da função *stem()* (linha 14 do quadro 6.2). Essa etapa é a responsável pela remoção das marcas de plural (*s*), passado (*ed*) e gerúndio (*ing*) dos termos avaliados.

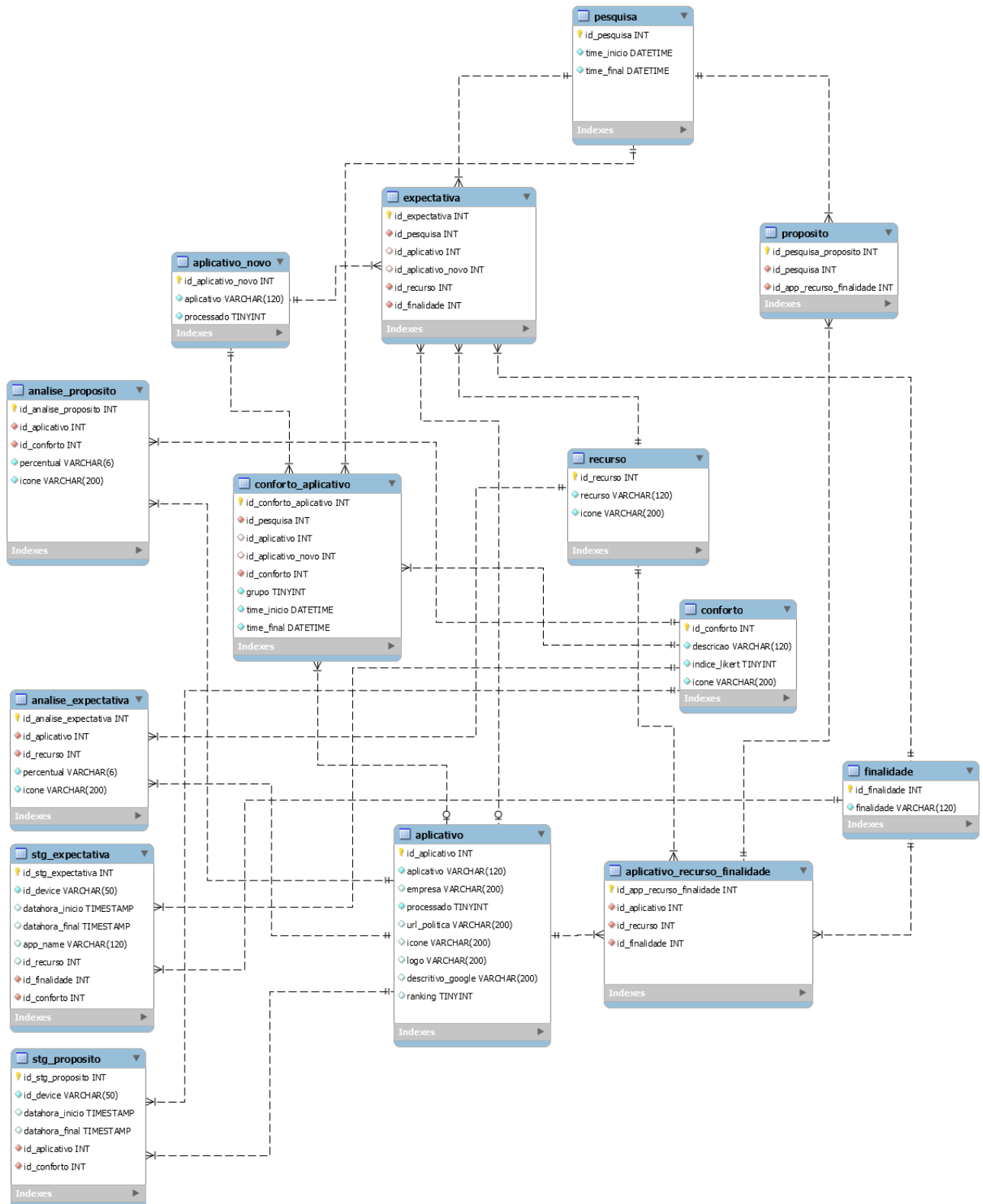
6.3.2 Armazenamento de dados

Todas as análises feitas pelo módulo de comportamento das aplicações e as respostas dos participantes da *crowdsourcing* foram armazenadas em uma tecnologia de banco de dados, instalado no servidor da aplicação.

Adotou-se como tecnologia de banco de dados para essa etapa a ferramenta MySQL. O modelo relacional, que mostra as tabelas do banco de dados e suas interconexões, foi desenvolvido de forma a suportar a inclusão de novos aplicativos ao serviço *WePrivacy*.

A figura 6.2 apresenta o modelo relacional desenvolvido. O modelo físico, relacionado à implantação dos objetos de banco de dados, pode ser consultado no Apêndice B.

Figura 6.2 - Modelo Relacional do banco de dados do serviço WePrivacy



O modelo é composto ao todo por 14 tabelas, dentre elas as tabelas *stg_expectativa* e *stg_propósito*, que são temporárias, utilizadas para armazenar as

respostas sem tratamento e os dados sem processamento fornecidos pelos participantes da *crowdsourcing*.

6.3.3 Processamento de resultados

As informações geradas pelo *PPMark* referentes ao comportamento real do aplicativo e as respostas dos participantes da *crowdsourcing* são analisadas e processadas neste módulo no servidor do *WePrivacy*.

As análises apresentadas na interface do aplicativo *WePrivacy* são o resultado de um processamento feito com o auxílio de ferramentas de análise de dados. Utilizou-se para isso a ferramenta *Pentaho Data Integrator* (PDI), também conhecida como *Kettle*.

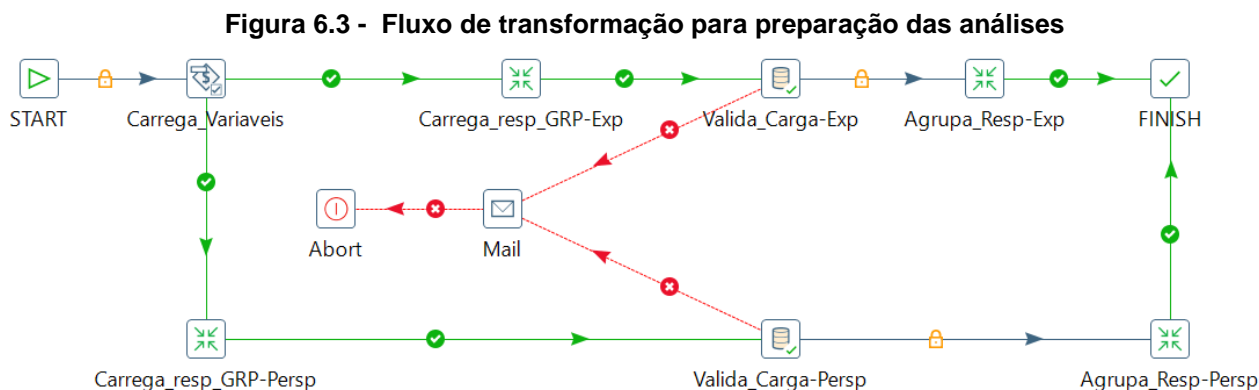
O *Kettle* é um *software open source* (código aberto) desenvolvido na linguagem java com o propósito de auxiliar os processos de análises e integração de dados, sendo muito utilizado como mecanismo de preparação de dados para a geração de relatórios de inteligência corporativa (PENTAHO, 2018).

O uso da ferramenta possibilitou a execução e a automatização do processo de carga dos dados das tabelas temporárias *stg_proposito* e *stg_expectativa* para as tabelas definitivas do modelo relacional.

O *Kettle* apresenta uma abordagem de funcionamento denominada fluxo de trabalho (*workflow*). O *workflow* está relacionado ao fluxo que o dado deve percorrer, passando por todas as etapas de transformação (*transforms*) de dados configuradas.

As transformações são as etapas que manipulam efetivamente os dados dentro de um fluxo de trabalho. São as responsáveis, em nosso contexto, por normalizar as respostas dos participantes da *crowdsourcing*, agregar e calcular os resultados para as análises apresentadas no *WePrivacy*.

A figura 6.3 ilustra o fluxo de trabalho denominado *j_weprivacy_analysis*, composto de 11 etapas de transformação. As transformações *Carrega_resp-GRP-Exp* e *Carrega_resp-GRP-Persp* são as responsáveis pelo carregamento dos dados das tabelas temporárias *stg_expectativa* e *stg_proposito*, respectivamente, para as tabelas definitivas do modelo relacional.



Após a etapa de carga dos dados, é feita uma validação (*Valida_Carga*) para ambos os grupos pesquisados (expectativa de coleta e perspectiva de conforto), a fim de verificar se a etapa anterior foi concluída com sucesso. Em caso de erro, um e-mail é enviado e o processo como um todo é cancelado pela etapa *Abort*.

As transformações *Agrupa_Resp-Exp* e *Agrupo_Resp-Persp* são as responsáveis, respectivamente, pelo agrupamento e processamento dos resultados mostrados na interface de análises do *WePrivacy*.

6.4 Interfaces do aplicativo *WePrivacy*

A prova de conceito para a abordagem de privacidade proposta neste trabalho, que se refere a considerar a expectativa das pessoas para apoiar outros usuários interessados na tomada de decisão de confiança no uso de aplicativos móveis, passa pela construção de aplicativo móvel que pudesse mostrar a viabilidade da implantação desse conceito.

O protótipo do *WePrivacy*, até o momento da escrita desse trabalho, está disponível apenas para versões do sistema operacional Android e pode ser encontrado na loja oficial da Google (Google Play) para *download*³ e instalação.

O protótipo proposto neste trabalho traz contribuições no sentido de incorporar as opiniões de outras pessoas no auxílio à tomada de melhores decisões de confiança em relação ao uso dos aplicativos móveis.

³ Disponível igualmente em: <http://mobile.privacidade.info>

O *WePrivacy* apresenta os equívocos mais comuns dos usuários em relação ao comportamento de um aplicativo móvel. Isso possibilita aos usuários interessados, ajustar seus modelos mentais em relação à privacidade de forma a reduzir a distância entre a expectativa e a realidade em relação ao comportamento de privacidade do aplicativo móvel.

Nosso projeto foca na apresentação das expectativas e dos efeitos do acesso a recursos dos dispositivos (nível de conforto) como sendo pontos-chaves que queremos transmitir aos usuários interessados no serviço.

Pesquisas anteriores, como em Felt et al. (2012) e Kelley et al. (2012), já discutiram diversos problemas de usabilidade e cognição apresentados pelos mecanismos de permissão, como, por exemplo, a tela de permissão do Android. Dentre os mais relevantes destacam-se: a formulação de uma longa lista de informações de permissão contendo muitos jargões técnicos para usuários leigos, poucas explicações sobre o potencial risco à privacidade e a fadiga que uma longa lista de informações de permissão pode ocasionar nos usuários.








Com esses problemas em mente, além dos dois pontos-chaves que já apresentamos, foram propostos alguns princípios para o protótipo do *WePrivacy* de forma a mitigar os problemas apresentados por Felt et al. (2012) e Kelley et al. (2012), sendo eles:

- O uso de termos simples para descrever os recursos relevantes. Por exemplo: em vez de usar termos como “IMEI (*International Mobile Equipment Identity*)” ou “ICCID (*Integrated Circuit Card Identifier*)”, optou-se simplesmente pelo uso de “Identificação do dispositivo móvel”, uma vez que são termos usados para identificar algum aspecto do dispositivo do usuário.
- Priorizar os recursos que tenham maior impacto sobre os sentimentos dos usuários, ou seja, sobre seus níveis de conforto.
- Destacar os pontos divergentes entre a expectativa e a realidade no que se refere ao comportamento do aplicativo. Foram utilizados para esta finalidade ícones para marcar e destacar esses cenários de divergências, como ilustrado no quadro 6.4.

Os ícones do tipo *smiles* foram usados para representar o consenso majoritário dos participantes do grupo perspectiva de conforto. Os demais ícones, do tipo geral, foram usados para representar o consenso dos participantes do grupo expectativa/

propósito, no qual foram analisados para cada um dos quatro recursos estudados (Lista de Contato, Localização de Rede, Localização Aproximada – GPS – e ID Dispositivo).

Quadro 6.4 - Descrição dos ícones de representação

Ícones	Tipo	Categoria de Uso	Descrição
	Smile	Perspectiva de Conforto	Aspecto <i> muito negativo</i> , associado a cenários nos quais 75% ou mais das pessoas que opinaram se sentiram <i> muito desconfortáveis</i> em relação ao comportamento de privacidade apresentado pelo aplicativo móvel.
	Smile	Perspectiva de Conforto	Aspecto <i> negativo</i> , associado à faixa de 50% a 74% das pessoas que opinaram e se sentiram <i> desconfortáveis</i> em relação ao comportamento de privacidade apresentado pelo aplicativo móvel.
	Smile	Perspectiva de Conforto	Aspecto <i> positivo</i> , associado à faixa de 50% a 74% das pessoas que opinaram e se sentiram <i> confortáveis</i> em relação ao comportamento de privacidade apresentado pelo aplicativo móvel.
	Smile	Perspectiva de Conforto	Aspecto <i> muito positivo</i> , associado a cenários nos quais 75% das pessoas que opinaram e se sentiram <i> muito confortáveis</i> em relação ao comportamento de privacidade apresentado pelo aplicativo móvel.
	Geral	Expectativa / Propósito	Aspecto <i> negativo</i> , associado à expectativa dos participantes. Representa cenários nos quais 33% ou menos dos usuários que opinaram tiveram a expectativa correta quanto ao comportamento do aplicativo analisado.
	Geral	Expectativa / Propósito	Aspecto <i> neutro</i> , associado à expectativa dos participantes. Refere-se a faixa de 34% a 65% dos usuários que opinaram e tiveram a expectativa correta quanto ao comportamento do aplicativo.
	Geral	Expectativa / Propósito	Aspecto <i> positivo</i> , associado à expectativa dos participantes. Representa cenários nos quais 66% ou mais dos usuários que opinaram, tiveram a expectativa correta quanto ao comportamento do aplicativo analisado.

As análises em ambos os grupos (expectativa de coleta/propósito e perspectiva de conforto) priorizam os aspectos negativos, ou seja, cenários que levam os usuários a se sentirem mais desconfortáveis.

A figura 6.4b ilustra a interface de análise do *WePrivacy*. Na sua parte inferior, denominada *estudos*, o sistema destaca, usando os ícones apresentados no quadro 6.4, os pontos de divergências entre expectativa e realidade.

Optou-se também por exibir no *WePrivacy*, na interface de análise, a lista dos recursos acessados e para qual propósito são utilizados (figura 6.4b). Isso possibilita aos usuários interessados conhecer o comportamento real do aplicativo e compará-lo às expectativas dos outros participantes.

A figura 6.4 apresenta dois exemplos de telas do projeto *WePrivacy*. A figura 6.4a refere-se à tela inicial do *WePrivacy*. Ao selecionar um dos aplicativos listados nesta tela, são exibidos o detalhamento e as análises feitas para o aplicativo selecionado. A figura 6.4b apresenta o detalhamento das análises para o aplicativo *Whatsapp*.

Figura 6.4 - Telas iniciais e de análise do *WePrivacy*



Nível de conforto, exibido na interface de análise, é um indicador criado que representa média aritmética simples dos níveis de conforto pontuados na pesquisa

para ambos os grupos. Utilizamos uma escala de Likert de quatro pontos, variando de -2, para muito desconfortável, a 2, para muito confortável.

A criação desse indicador está em consonância com o estudo apresentado por Kelley et al. (2012), que relata a influência de indicadores, como, por exemplo, *ranking* de qualidade, no suporte ao usuário para tomada de decisões sobre a qualidade do aplicativo. Neste estudo, adota-se como indicador a variável nível de conforto, já utilizada nos questionários de pesquisa com os usuários.

A possibilidade de expansão do serviço para outros aplicativos, agregando opiniões de outros usuários, foi uma preocupação no projeto do *WePrivacy*. Existem basicamente duas formas de contribuir com o serviço: por meio do botão participar, presente na tela inicial, ilustrado na figura 6.4a, ou por meio do botão opinar/contribuir, presente na tela de análises, ilustrada na figura 6.4b.

Ao participar pela tela inicial, o *WePrivacy* categorizará o participante dentro do grupo de expectativa e selecionará aleatoriamente um dos aplicativos móveis que o participante tenha instalado em seu dispositivo para responder ao questionário.

Por meio dessa etapa é possível a expansão do serviço, incorporando outros aplicativos além dos iniciais, já inclusos. A seleção do aplicativo a partir da opção de participar pela tela inicial segue três critérios:

- I. O aplicativo selecionado não deve ter as análises (figura 6.4b) visualizadas pelo usuário, pois neste grupo (expectativa de coleta/uso) o participante não é informado sobre o comportamento do aplicativo. Esse requisito é necessário para evitar a tendência no resultado da pesquisa e por consequência na contribuição do participante para o serviço.

O *WePrivacy* armazena a lista de aplicativos cujas análises o participante já visualizou, e, por consequência, aplicativos desta lista não são selecionados.

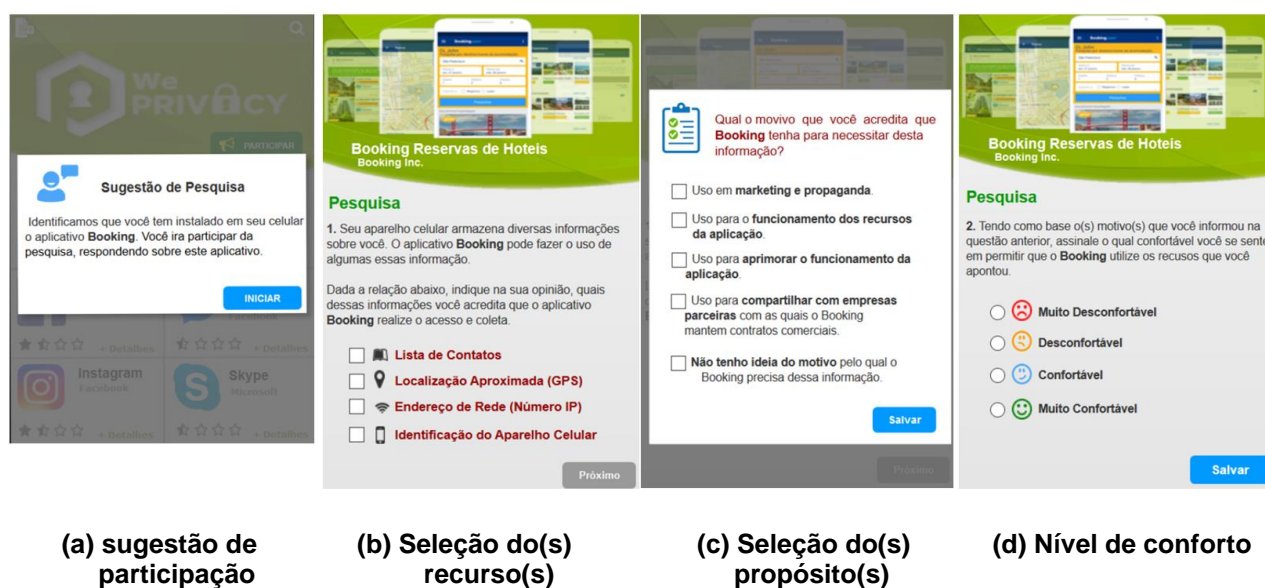
- II. O serviço busca selecionar os aplicativos de forma a obter um número de resposta igualitária entre os grupos de coleta/uso e perspectiva de conforto, respeitando o primeiro critério, ou seja, sempre que possível o *WePrivacy* busca deixar todos os aplicativos com o mesmo número de respostas e participantes. No caso de algum aplicativo já presente no serviço estar abaixo dos demais

quanto ao número de participantes, este aplicativo será o selecionado para o usuário fornecer sua contribuição.

- III. No cenário no qual todos os aplicativos já estejam com um número igualitário de respostas, um novo aplicativo é selecionado a partir do dispositivo do usuário para compor o serviço. O usuário então responde sobre suas expectativas em relação ao comportamento que ele acredita que seja adotado pelo aplicativo selecionado.

A figura 6.5 apresenta quatro exemplos de interfaces do *WePrivacy* referentes ao fluxo da etapa de participação. Este exemplo, ilustra um cenário no qual o usuário foi categorizado no grupo expectativa de coleta/uso (botão *participar* a partir da tela inicial figura 6.4a).

Figura 6.5 - Telas do fluxo de participação para o grupo expectativa de coleta/ uso.



No exemplo ilustrado, o sistema selecionou o aplicativo móvel *Booking Reservas de Hotéis* e então, a partir da seleção de um recurso que o usuário acredite que o aplicativo efetue o acesso (figura 6.5b), é exibida na tela uma lista para que o participante informe o motivo que ele acredita que o aplicativo tenha para acessar a informação solicitada (figura 6.5c). Por fim, o usuário é questionado sobre o seu nível

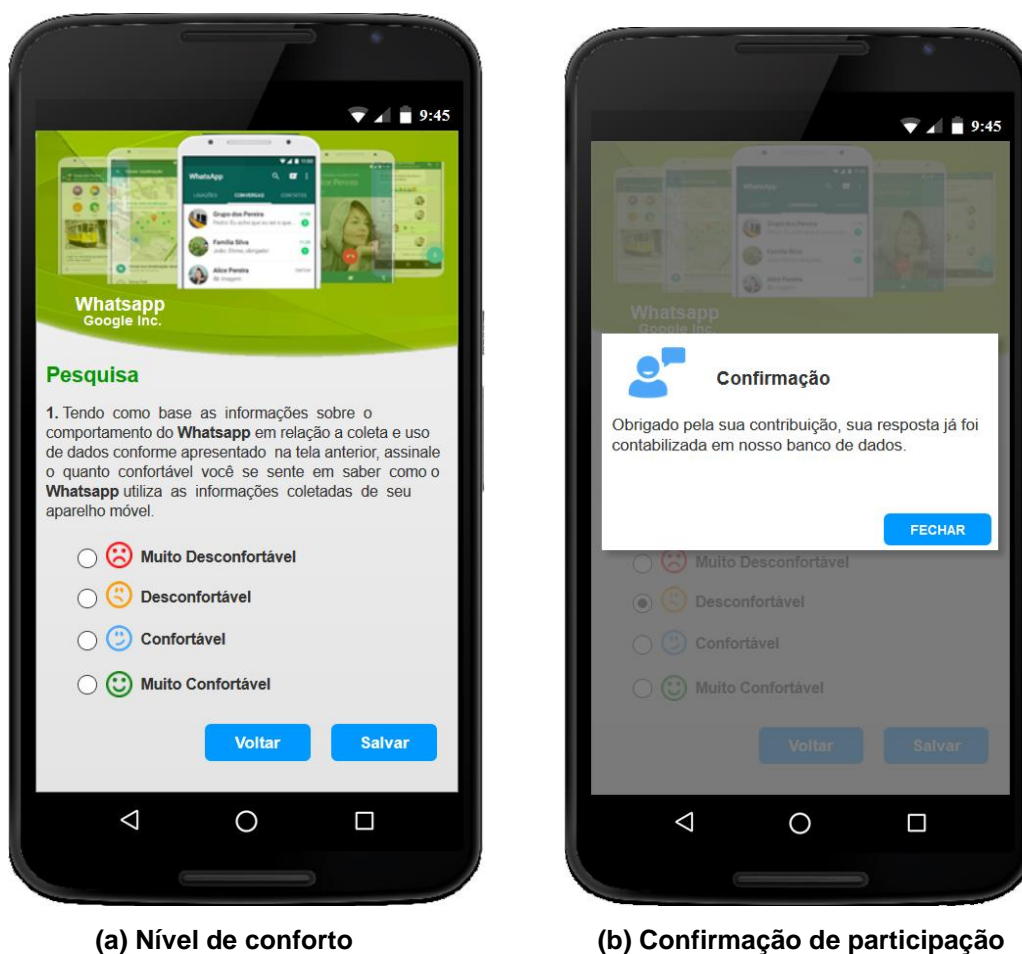
de conforto em relação ao comportamento que ele acredita que o aplicativo avaliado tenha (figura 6.5d).

Ao contribuir com o serviço a partir da interface de análises (figura 6.4b), o *WePrivacy* categoriza o participante no grupo perspectiva de conforto. Nesta categoria os participantes são informados pela própria tela do *WePrivacy* sobre quais recursos o aplicativo analisado coleta e para que propósito serve a informação obtida. Os usuários respondem, então, o quão confortáveis se sentem em saber sobre esse comportamento do aplicativo.

A avaliação das respostas dos usuários do grupo perspectiva de conforto dá origem às análises categorizadas pelos ícones do tipo *smiles*, que podem ser observados no quadro 6.4 e na tela de análises do *WePrivacy* (figura 6.4b).

A figura 6.6 apresenta as duas telas de interação que são exibidas aos usuários que contribuírem com o serviço por meio do botão opinar/contribuir na interface de análises (figura 6.4b).

Figura 6.6 - Telas do fluxo de participação para o grupo perspectiva de conforto



Conforme foi ilustrado pela figura 6.6a, o questionário apresenta apenas a questão referente ao sentimento do usuário (nível de conforto) em relação ao aplicativo móvel sobre o qual ele escolheu opinar, neste caso, o aplicativo *Whatsapp*. Após, como observado na figura 6.6b, uma mensagem de confirmação é exibida para informar o final do processo.

CAPÍTULO 7

AVALIAÇÃO DO *WEPRIVACY*

Este capítulo traz a avaliação do protótipo WePrivacy proposta por este trabalho. A seção 7.1 apresenta as considerações iniciais referentes a um segundo estudo desenvolvido com os usuários tendo como objetivo a validação da abordagem de privacidade do WePrivacy. A seção 7.2 discorre sobre as etapas adotadas na condução deste estudo e por fim a seção 7.3 apresenta e discute os resultados encontrados.

7.1 Considerações iniciais

No capítulo anterior, apresentamos e discutimos o protótipo desenvolvido para implementar a prova do conceito de privacidade usado neste trabalho. Neste capítulo são apresentados o planejamento, a execução e os resultados da avaliação do protótipo do *WePrivacy* proposto nesta dissertação.

Utilizou-se como elemento de avaliação do *WePrivacy* um segundo estudo, conduzido com a mesma população acadêmica apresentada anteriormente, no capítulo 4. Selecionamos uma segunda amostra, constituída de 118 participantes, na qual eles foram convidados a participar de uma pesquisa que lhes solicitou interagir com dois mecanismos de proteção à privacidade, sendo um deles o *WePrivacy*.

Cada participante interagiu com os dois mecanismos de proteção à privacidade, observando as informações referentes a três aplicativos móveis previamente selecionados a partir da pesquisa primária apresentada e discutida nos capítulos 4 e 5, respectivamente.

7.2 Condução da avaliação

Para a avaliação do *WePrivacy* optou-se pela seleção apenas da interface de análises, como ilustrado na figura 7.1. Como essa interface representa a abordagem de privacidade discutida neste trabalho, ela foi utilizada na comparação da interface do mecanismo de notificação de privacidade do *Android*, figura 7.2. Queríamos observar se a interface do *WePrivacy*, que considera a expectativa das pessoas, forneceria um suporte melhor aos participantes em relação às suas decisões de confiança e privacidade no que se refere ao comportamento dos aplicativos móveis.

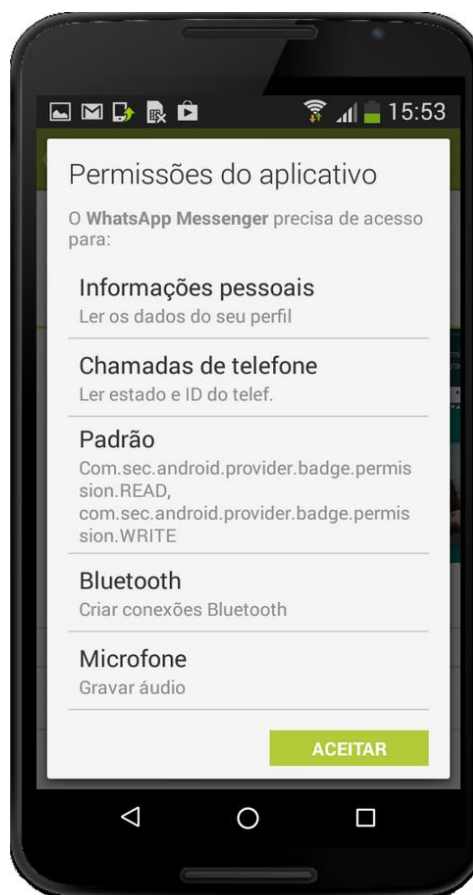
Figura 7.1 - Tela de exibição das análises do *WePrivacy*



Foram utilizados os dados obtidos no estudo primário para alimentar as interfaces do *Android* e *WePrivacy* para os três aplicativos móveis que mais geraram divergência entre a expectativa e o comportamento real, apresentados no estudo primário. Os aplicativos selecionados foram *Snapchat*, *UC Browser* e *YouTube*.

Optou-se pela seleção de apenas três aplicativos para evitar fadiga nos participantes devido ao número de questões a serem respondidas.

Figura 7.2 - Tela de privacidade para instalação de aplicativos em sistema *Android* - Adaptado de Felt et al. (2012)



Os dados foram coletados por meio de um questionário composto de três sessões: (I) Termo de Consentimento Livre e Esclarecido (TCLE), no qual os participantes formalizavam o aceite em participar da pesquisa; (II) Questões demográficas como idade, sexo e o curso ao qual pertence; (III) Afirmações utilizadas na avaliação das interfaces.

Utilizamos uma escala de Likert de quatro pontos para avaliar a percepção dos usuários em relação aos elementos apresentados. Adotamos valores de 1 para os itens de menor percepção e 4 para os itens de maior percepção, segundo as respostas dos participantes.

O quadro 7.1 ilustra o formulário utilizado na avaliação do mecanismo *WePrivacy*, na etapa (III).

Antes de responder às afirmações da seção (III), os participantes foram convidados a observar e interagir com as interfaces avaliadas (*Android* e *WePrivacy*).

Quadro 7.1 - Formulário usado na avaliação de interface

Aspectos de Usabilidade			
Nº	Itens de avaliação		Escala Likert
1	Leitura de caracteres na tela	Difícil	○ ○ ○ ○ 1 2 3 4 FÁCIL
2	Disposição das informações na tela	Mal distribuída	○ ○ ○ ○ 1 2 3 4 Bem distribuída
3	Volume de informações apresentadas	Inadequado	○ ○ ○ ○ 1 2 3 4 Adequado
4	Termos utilizados são compreensíveis	Discordo plenamente	○ ○ ○ ○ 1 2 3 4 Concordo plenamente
5	Facilidade no uso da aplicação	Difícil	○ ○ ○ ○ 1 2 3 4 FÁCIL
Aspectos de Efetividade			
6	Capacidade de identificar quais informações são coletadas	Difícil	○ ○ ○ ○ 1 2 3 4 FÁCIL
7	Capacidade de identificar os objetivos da coleta da informação	Difícil	○ ○ ○ ○ 1 2 3 4 FÁCIL
8	A interface ajuda na tomada de decisão sobre instalar ou não um determinado aplicativo móvel.	Discordo plenamente	○ ○ ○ ○ 1 2 3 4 Concordo plenamente
9	Considero a interface uma ferramenta útil para uso no cotidiano.	Discordo plenamente	○ ○ ○ ○ 1 2 3 4 Concordo plenamente

Cada participante observou e interagiu com as duas interfaces nos três aplicativos avaliados, totalizando seis telas entre o *Android* e o *WePrivacy*. Essa interação se deu por meio dos computadores do laboratório da IES, onde a pesquisa foi desenvolvida.

Ao longo da aplicação do questionário, os participantes poderiam voltar a consultar as interfaces novamente, se assim o desejassem. Cada participante recebeu dois formulários da seção (III), um relacionado à interface do *Android* e outro idêntico, referente à interface do *WePrivacy*.

7.3 Resultados da avaliação

Foram obtidos ao todo 236 questionários respondidos, sendo que 12, ou 4,68%, foram desconsiderados por não terem sido aprovados no quesito qualidade.

O quesito qualidade foi avaliado levando-se em conta três fatores: o primeiro relacionado ao cenário no qual o participante assinala o mesmo valor para todas as afirmações. O segundo fator está associando ao cenário no qual o participante assinala para uma mesma afirmação mais de um valor resposta. Por fim, o terceiro fator de qualidade refere-se às situações nas quais o questionário continha algumas das afirmações sem nenhuma resposta assinalada, ou seja, afirmações em branco.

Os questionários que não se enquadraram nos critérios de qualidade foram descartados. Os participantes foram informados no início da pesquisa sobre esses critérios.

O estudo buscou avaliar as interfaces sob duas perspectivas. A primeira foi a usabilidade, medida com o recurso à média aritmética e do desvio padrão das respostas assinaladas pelos participantes referentes às cinco primeiras afirmações do questionário (afirmações do grupo de usabilidade). A segunda perspectiva estudada é a efetividade. Sua avaliação se deu por meio da análise das respostas das quatro últimas afirmações do formulário (6 - 9), como ilustrado pelo quadro 7.1.

As análises entre as duas perspectivas avaliadas para cada interface estão detalhadas nas tabelas 7.1 e 7.2, respectivamente. A coluna Pts (pontos) refere-se à soma dos valores obtidos na escala Likert para a afirmação correspondente do formulário, relacionada a todos os aplicativos avaliados.

De forma semelhante, a coluna Avg apresenta a média aritmética dessas respostas e a coluna Std apresenta, por fim, o cálculo do desvio padrão para cada cenário avaliado.

O desvio padrão, segundo Morettin e Bussab (2010), é uma medida de dispersão e mede a variação dos dados em torno do valor médio calculado na amostra analisada.

Um valor de desvio padrão muito baixo indica que os pontos dos dados tendem a estar próximos do valor médio. Isso indica uma característica de amostras com comportamento muito homogêneo, com pouca variabilidade nas respostas. Por outro lado, um valor muito alto de desvio padrão sinaliza comportamentos heterogêneos,

indica alta variabilidade nas respostas e, por conseguinte, uma maior dispersão dos dados em torno da média amostral (MORETTIN; BUSSAB, 2010).

Tabela 7.1 - Comparação de resultados pela perspectiva de usabilidade

Perspectiva de Usabilidade						
	WePrivacy			Android		
	Pts	Avg	Std	Pts	Avg	Std
Afirmção 1	273	2.44	1.13	254	2.27	1.09
Afirmção 2	297	2.65	1.11	273	2.44	1.07
Afirmção 3	289	2.58	1.07	260	2.32	1.09
Afirmção 4	294	2.63	1.11	258	2.30	1.14
Afirmção 5	236	2.11	0.97	284	2.54	1.16

De um modo geral, os participantes acreditam que a interface de análise do *WePrivacy* oferece usabilidade superior quando comparada ao mecanismo padrão de permissão do *Android*. Contudo, a avaliação sobre a facilidade do uso da interface (Afirmção 5) obteve melhor pontuação na interface do *Android*. Tal achado pode ter relação com o fato da interface ser somente um mecanismo de notificação simples, que apenas serve ao propósito de notificar qual recurso será acessado. Não existem, nesse mecanismo, maiores detalhes sobre esse acesso, como, por exemplo, o motivo de o aplicativo acessar o recurso. A interação com a interface é mínima, limitando-se apenas a um botão, como ilustrado na figura 7.2. Essa simplicidade pode ser o fator que gerou nos participantes a sensação de facilidade de uso.

Tabela 7.2 - Comparação de resultados pela perspectiva de efetividade

Perspectiva de Efetividade						
	WePrivacy			Android		
	Pts	Avg	Std	Pts	Avg	Std
Afirmção 6	317	2.83	0.98	282	2.52	1.09
Afirmção 7	339	3.03	0.89	N/A*	N/A*	N/A*
Afirmção 8	288	2.57	1.22	270	2.41	1.14
Afirmção 9	286	2.55	1.13	280	2.50	1.14

* N/A - Não aplicado, pois a interface não tem suporte para informar o objetivo da coleta.

Nos aspectos relacionados à efetividade, os participantes também acreditam que a interface do *WePrivacy* é superior em comparação ao mecanismo padrão de permissão do *Android*.

A tabela 7.2 apresenta o detalhamento dessa análise, exibindo o cômputo dos resultados para o grupo de afirmações referentes à avaliação da efetividade (6 - 9).

A avaliação promovida pela Afirmação 7, que se refere à capacidade da interface de informar o motivo do acesso a determinado recurso do aplicativo, não foi considerada para análise no mecanismo de controle da privacidade no *Android*. A interface não oferece suporte a nenhuma informação sobre o propósito que determinado aplicativo teria para acessar o recurso solicitado.

Optou-se por manter essa análise na interface *WePrivacy*, pois desejava-se identificar se as pessoas conseguiam, por meio da observação e interação, identificar qual o objetivo dos aplicativos analisados em acessar os recursos apresentados.

A tabela 7.3 apresenta um resumo das análises anteriores, em ambas perspectivas (usabilidade e efetividade).

Tabela 7.3 - Resumo das análises de usabilidade e efetividade

Análises por perspectivas						
Interface	Usabilidade			Efetividade		
	Pts	Avg	Std	Pts	Avg	Std
WePrivacy	1389	2.48	1.10	1230	2.75	1.08
Android	1329	2.37	1.11	832	2.48	1.13

Os participantes consideraram a interface do *WePrivacy* superior em relação ao mecanismo de permissão do *Android*. A Afirmação 7, discutida anteriormente, não foi considerada para o cálculo da efetividade na interface do *Android*.

A análise da usabilidade foi obtida com a avaliação de todas as afirmações (1 - 5) apresentadas na tabela 7.1. De forma semelhante, a análise da efetividade considerou o grupo de afirmações (6 - 9), excluindo a Afirmação 7, como informado anteriormente.

CAPÍTULO 8

CONCLUSÃO E TRABALHOS FUTUROS

Este capítulo apresenta as conclusões deste trabalho, bem como as contribuições de pesquisa e as propostas de trabalhos futuros.

8.1 Considerações iniciais

Este trabalho apresentou uma abordagem de privacidade, considerando a expectativa das pessoas para apoiar a tomada de decisão sobre o uso ou não de determinado aplicativo móvel.

A prova de conceito para esta abordagem foi desenvolvida com a implantação de um protótipo de serviço denominado *WePrivacy*, que fornece para os usuários interessados, por meio de um aplicativo móvel, informações de privacidade, sobre o comportamento dos aplicativos móveis e sobre como as pessoas se sentiram frente a este comportamento.

O nível de conforto das pessoas fora inicialmente obtidos com a aplicação de um estudo do tipo *survey* (*pesquisa*), no qual um grupo de pessoas de determinada comunidade acadêmica relatou suas expectativas em relação ao comportamento do aplicativo avaliado. Após a implantação do serviço *WePrivacy*, os próprios usuários interessados podem contribuir com o serviço, fornecendo suas opiniões em relação a novos aplicativos ou aos aplicativos já avaliados. Isso possibilita a expansão da proposta do serviço a outros domínios de aplicativos e a comunidades de pessoas para além da consultada neste trabalho.

A longo prazo, a expectativa em relação ao *WePrivacy* é se tornar um serviço de apoio à avaliação de privacidade escalável para vários outros tipos de aplicações móveis, combinando técnicas de análise de aplicações com as opiniões/percepções

das pessoas (*crowdsourcing*).

As técnicas de análises são destinadas à obtenção do comportamento real da aplicação. Já a *crowdsourcing* visa a capturar as percepções e expectativas das pessoas de modo colaborativo em relação ao comportamento do aplicativo móvel. Na proposta apresentada pelo *WePrivacy*, o *crowdsourcing* é obtido com a possibilidade de os usuários interessados contribuírem com o serviço, fornecendo suas opiniões e com isto ajudando a expandir a base inicial do serviço.

8.2 Contribuições de pesquisa

Uma observação importante contida em nosso trabalho está relacionada aos sentimentos que os participantes manifestaram quando participaram da pesquisa. De maneira geral, os participantes se sentiram mais confortáveis quando foram informados os motivos do acesso a determinado recurso pelos aplicativos.

Os participantes foram expostos a casos nos quais uma funcionalidade não era usada para a função principal do aplicativo, como, por exemplo, o Facebook acessando o endereço de rede (IP) para fins de compartilhamento com empresas parceiras. Isso conflitou com algumas expectativas iniciais dos participantes (grupo de expectativas). Dentre os usuários que tinham a expectativa desse acesso, a maioria achou ser para fins de publicidade ou formação de perfil de uso (rastreamento). Em decorrência disso, atribuíram o nível de conforto muito menor em comparação ao grupo de participantes que teve revelado os motivos do acesso aos recursos dos dispositivos móveis.

Observaram-se também alguns casos (por exemplo, o aplicativo 360 Mobile Security) em que os participantes tiveram expectativas corretas sobre o aplicativo acessar a lista de contatos para prover o serviço principal. Contudo, o nível de conforto foi menor quando comparado com a condição do grupo de participantes que não foi informado sobre o real propósito do acesso ao recurso. Essa característica nos sugere que, para esses casos, ao lidar com as incertezas, os usuários tendem a se preocupar mais com algo que possa comprometer a sua privacidade.

Atualmente, a tela de permissão do *Android*, exibida quando se deseja instalar algum aplicativo móvel desta plataforma, não contém quaisquer explicações sobre o

motivo do recurso ser acessado (semelhante à condição de expectativa que apresentamos neste estudo). Como discutimos anteriormente, informar os motivos pode ajudar os usuários a tomarem melhores decisões de confiança e aprimorar até certo ponto o seu nível de conforto.

Uma abordagem possível é utilizar de *feedbacks* dos usuários para compor um indicador que apresente esse nível de conforto em relação à solicitação da permissão de acesso. Tal abordagem é implementada pela proposta do *WePrivacy* que apresentamos e discutimos neste trabalho.

Observou-se que os aplicativos que utilizam dados dos acessos aos recursos dos dispositivos móveis para fins de publicidade são uma preocupação majoritária entre os participantes. Para todos os quatro tipos de recursos que avaliamos neste estudo, os usuários se sentiram mais desconfortáveis quando esse acesso tinha como objetivo a publicidade.

Entendemos que muitas empresas dependem de publicidade para monetizar suas aplicações. No entanto, para melhorar a experiência dos usuários dessas aplicações, deveriam informá-los sobre como e porque suas informações são utilizadas. Elas podem ajudar os usuários na questão da privacidade sem comprometer a qualidade do serviço de publicidade. Outras maneiras potenciais incluem o uso de tecnologias para esconder o dado real do usuário, por meio, por exemplo, de substituições, uso de *hash* ou criptografias.

Por fim, uma contribuição que destacamos neste trabalho é a possibilidade de incluir as expectativas das pessoas na avaliação da privacidade. Apresentou-se uma abordagem de como capturar essas expectativas e identificar nelas os pontos fortes - quando em consonância com o comportamento do app - e fracos - quando em discordância com o comportamento do app.

8.3 Conclusões

Diversos trabalhos encontrados na literatura especializada apresentam contribuições para a área de privacidade e segurança, principalmente com foco em fornecer ferramentas de análise automatizadas no sentido de ajudar os usuários a controlar a sua privacidade. No entanto, não se encontra nesses trabalhos nenhum

mecanismo capaz de distinguir se o acesso a determinado recurso móvel é necessário, ou capaz de apontar qual a sensação que essa incerteza causa aos usuários, pois têm de permitir o acesso aos recursos sem saber ao certo para quais propósitos eles serão utilizados.

Nossa pesquisa apresentou uma maneira de abordar a privacidade em aplicações móveis, considerando a expectativa das pessoas. Em nosso contexto, exploramos os sentimentos de cada pessoa sobre suas expectativas de uso e de acesso a determinado recurso móvel.

Nossos resultados sugerem que tanto a expectativa de acesso quanto a de propósito têm grande impacto sobre os níveis de conforto dos usuários e podem afetar suas decisões de confiança. Outra importante observação é que, quando os usuários são informados corretamente sobre o acesso a determinado recurso e o propósito deste acesso é suficientemente claro, torna-se possível atenuar o desconforto de privacidade dos usuários.

Tendo como referência as informações levantadas nesta pesquisa, propôs-se um serviço denominado *WePrivacy*, que apresenta aos usuários interessados o comportamento real de um aplicativo móvel e as opiniões de outras pessoas sobre o comportamento daquele aplicativo, destacando as situações de divergência entre a expectativa e o comportamento real do aplicativo.

A nossa interface foi considerada, na visão dos participantes, superior em diversos aspectos em relação ao mecanismo de permissão do sistema *Android*. Dentre esses aspectos estão o fornecimento de informações mais pertinentes para ajudar os usuários a tomar melhores decisões de privacidade e confiança relacionadas ao uso de determinado aplicativo móvel.

8.4 Trabalhos futuros

Existem diversas vertentes que possibilitam expandir a pesquisa atual. Como propostas para trabalhos futuros, pode-se citar, por exemplo, o aperfeiçoamento e o desenvolvimento das funcionalidades atribuídas ao *WePrivacy*, como:

- Expansão para análise considerando outros grupos de recursos, para além dos quatro estudados neste trabalho;

- A portabilidade do código para um modelo de *Cloud Service* (Serviço de nuvem), que tornaria desnecessário instalar o aplicativo do serviço;
- O aprofundamento de técnicas de inteligência artificial, mais especificamente em análises de textos, a fim de melhorar a capacidade do *WePrivacy* na compreensão das políticas de privacidade dos aplicativos avaliados e com isso ampliar a possibilidade da criação de novos grupos de propósitos em adição aos já existentes.
- Por fim, a possibilidade de o mecanismo *WePrivacy* incluir uma análise sobre o destino da informação coletada. Atualmente o serviço apenas considera o propósito. Agregar uma avaliação quanto ao destino permitiria verificar se determinada informação coletada é enviada à empresa que produziu o aplicativo e como seria a percepção do usuário em relação a esses novos comportamentos.

8.5 Publicação de resultados

Ao longo do desenvolvimento do trabalho de mestrado, os resultados e discussões aqui apresentados foram submetidos e aceitos para publicação em conferências acadêmicas, como o *Americas Conference on Information Systems* (AMCIS).

Pretende-se, ainda, a submissão do trabalho a um *journal* no qual os resultados e discussões possam ser apresentados de maneira mais pormenorizada.

REFERÊNCIAS

AGRE, P. E.; ROTENBERG, M. **Technology and privacy: The new landscape**. [S.l.]: Mit Press, 1998. 336 p. ISBN 978-02-62-51101-8.

AKINS, R. N. Measurement and Research Methodology. In: **American Educational Research Association - AERA Division D: Forum NJ Dept. of education**. 2002. Disponível em: <[http://www.aera.net/Division-D/Measurement-Research-Methodology - D](http://www.aera.net/Division-D/Measurement-Research-Methodology-D)>. Acesso em: 07 Jun. 2018.

ALEXANDRE, J. W. C. et al. Análise do número de categorias da escala de likert aplicada à gestão pela qualidade total através da teoria da resposta ao item. **Encontro Nacional De Engenharia De Produção**, v. 23, p. 1–20, 2003.

ANNIE, A. **App Annie 2015 Retrospective - Research & Analysis**. 2015. Disponível em: <<http://files.appannie.com.s3.amazonaws.com/reports/App-Annie-2015-Retrospective-EN-1.pdf>>. Acesso em: 07 Jun. 2018.

BARR, P. S.; STIMPERT, J. L.; HUFF, A. S. Cognitive change, strategic action, and organizational renewal. **Strategic management journal**, Wiley Online Library, v. 13, n. S1, p. 15–36, 1992.

BASTOS, V. M. **Ambiente de Descoberta de Conhecimento na Web para a Língua Portuguesa**. 133 p. Tese de Doutorado. Universidade Federal do Rio de Janeiro, 2006.

BHASKAR, P.; AHAMED, S. I. Privacy in pervasive computing and open issues. In: **ARES 2007. the second international conference on availability, reliability and security**. Washington, DC, USA: [s.n.], 2007. p. 147–154. ISBN 0-7695-2775-2.

BRASIL. **Lei n. 12.965, 23 de Abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil**. Brasília: Diário Oficial da União, 2014.

_____. **Ante Projeto de Lei. Dispõe sobre o tratamento de dados pessoais para a garantia do livre desenvolvimento da personalidade e da dignidade da pessoa natural**. 2015. Disponível em: <<http://www.justica.gov.br/noticias/mj-apresenta-nova-versao-do-anteprojeto-de-lei-de-protecao-de-dados-pessoais/apl.pdf>>. Acesso em: 07 Jun. 2018.

BRAVO-LILLO, C. et al. Bridging the Gap in Computer Security Warnings: A Mental Model Approach. **IEEE Security & Privacy Magazine**, v. 9, n. 2, p. 18–26, mar 2011. ISSN 1540-7993. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5669245>>.

CARBUNAR, B.; POTHARAJU, R. A Longitudinal Study of the Google App Market. In: **Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015 - ASONAM '15**. New York, USA: ACM Press, 2015. p. 242–249. ISBN 978-14-50-33854-7. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2808797.2808823>>.

CHAPMAN, J. A.; FERFOLJA, T. Fatal flaws: the acquisition of imperfect mental models and their use in hazardous situations. **Journal of Intellectual Capital**, MCB UP Ltd, v. 2, n. 4, p. 398–409, 2001.

CHERMACK, T. J. Mental models in decision making and implications for human resource development. **Advances in developing human resources**, Sage Publications, v. 5, n. 4, p. 408–422, 2003.

COCHRAN, W. G. **Sampling techniques**. 3. ed. Westlake Village: John Wiley & Sons, 1977. 428 p. (Wiley series in probability and mathematical statistics: Applied probability and statistics). ISBN 0-471-16240-X.

CRAIK, K. **The Nature of Explanation**. Cambridge University Press, 1967. 136 p. (Philosophy:science CAM). ISBN 978-05-21-09445-0.

DONEDA, D. **Da privacidade à proteção de dados pessoais**. Rio de Janeiro: Renovar, 2006. 439 p. ISBN 978-85-71-47562-5.

DUARTE, J. C. **O Algoritmo Boosting at Start e suas Aplicações**. 87 p. Tese de Doutorado. PUC - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, BR, 2009.

EBECKEN, N. F.; LOPES, M. C. S.; COSTA, M. C. **Mineração de textos. Sistemas inteligentes: fundamentos e aplicações**. São Carlos: Manole, p. 337–370, 2003.

ENCK, W. et al. Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. **ACM Transactions on Computer Systems - TOCS**, ACM, New York, NY, USA, v. 32, n. 2, p. 5:1–5:29, jun. 2014. ISSN 0734-2071. Disponível em: <<http://doi.acm.org/10.1145/2619091>>.

FELT, A. P. et al. Android permissions: User attention, comprehension, and behavior. In: **Proceedings of the eighth symposium on usable privacy and security**. New York, NY, USA: ACM, 2012. (SOUPS '12), p. 3:1–3:14. ISBN 978-14-50-1532-6. Disponível em: <<http://doi.acm.org/10.1145/2335356.2335360>>.

FTC, Federal Trade Commission. **About the FTC**. 2018. Disponível em: <<https://www.ftc.gov/about-ftc>>. Acesso em: 07 Jun. 2018.

FOX, C. A stop list for general text. **Special Interest Group on Information Retrieval - SIGIR**, ACM, New York, NY, USA, v. 24, n. 1-2, p. 19–21, set. 2003. ISSN 0163-5840. Disponível em: <<http://doi.acm.org/10.1145/378881.378888>>.

GARDNER, H. Mentas que mudam: **A arte e a ciência de mudar as nossas ideias e as dos outros**. Artmed, 2005. 229 p. ISBN 9788536304281.

GARSON, G. D. **Generalized Linear Models and Generalized Estimating Equations, from Statnotes: Topics in multivariate analysis**. 2011. Disponível em: <<http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm>>. Acesso em: 25 Abr. 2018.

GHANI, N.; SIDEK, Z. Controlling and disclosing your personal information. **WSEAS Transactions on Information Science and Applications**, v. 6, n. 3, p. 397–406, 2009. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-66349108406&partnerID=40&md5=5078090b72a4b0bbbb9bf258f09f5a64>>.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010. 184 p. ISBN 978-85-22-45823-3.

GOOGLE. **API Guides: App Manifest**. 2016. Disponível em: <<https://developer.android.com/guide/topics/manifest/manifest-intro.html>>. Acesso em: 07 Jun. 2018.

GOULART, G. D. **Dados Pessoais e Dados Sensíveis: a insuficiência da categorização**. 2015. Disponível em: <<http://direitoeti.com.br/artigos/dados-pessoais-e-dados-sensiveis-a-insuficiencia-da-categorizacao>>. Acesso em: 07 Jun. 2018.

HORNYACK, P. et al. These aren't the droids you're looking for: Retrofitting android to protect data from imperious applications. In: **Proceedings of the 18th acm conference on computer and communications security**. New York, NY, USA: ACM, 2011. (CCS '11), p. 639–652. ISBN 978-1-4503-0948-6. Disponível em: <<http://doi.acm.org/10.1145/2046707.2046780>>.

KELLEY, P. G. et al. A "nutrition label" for privacy. In: **Proceedings of the 5th symposium on usable privacy and security soup '09**. 2009. Disponível em <http://portal.acm.org/citation.cfm?doid=1572532.1572538>

_____. A conundrum of permissions: Installing applications on an android smartphone. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 7398 LNCS, p. 68–79, 2012. ISSN 03029743

KPCB, Kleiner Perkins Caufield Byers. **2015 Internet Trends**. 2015. Disponível em <<http://www.kpcb.com/file/2015-internet-trends-report>>. Acesso em: 07 Jun. 2018.

_____. **2017 Internet Trends**. 2017. Disponível em <<http://www.kpcb.com/file/2015-internet-trends-report>>. Acesso em: 07 Jun. 2018.

LANGHEINRICH, M. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. In: **International conference on Ubiquitous Computing**. Springer Berlin Heidelberg, 2001. p. 273-291.

LECHETA, R. R. **Web services RESTful: Aprenda a criar web services RESTful em Java na nuvem do Google**. Novtec, 2015. 432 p. ISBN 978-85-75-22454-0.

LIM, B. C.; KLEIN, K. J. Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. **Journal of Organizational Behavior**, v. 27, n. 4, p. 403–418, 2006.

LOVINS, J. B. **Development of a stemming algorithm**. [S.l.]: MIT Information Processing Group, Electronic Systems Laboratory Cambridge, 1968.

MANSILHA, S. **Comunicação Corporativa**. Clube de Autores, 2008. 213 p.

MCDONALD, A. M.; CRANOR, L. F. Cost of reading privacy policies, the. **Journal of Law and Policy for the Information Society**, v. 4, p. 543, 2008.

MICHAEL, J. **Privacy and human rights: An international and comparative study, with special reference to developments in information technology**. Dartmouth Pub Co. UNESCO, 1994.

MILLER, J.; SUMEETH, M.; SINGH, R. I. Evaluating the readability of privacy policies in mobile environments. **International Journal of Mobile Human Computer Interaction**, IGI Global, Hershey, PA, USA, v. 3, n. 1, p. 55–78, jan. 2011. ISSN 1942-390X. Disponível em: <<http://dx.doi.org/10.4018/jmhci.2011010104>>.

MOORE, D. **The Basic Practice of Statistics**. 5 ed. New York: Freeman, 2010. 730 p. ISBN 978-14-29-22426-0.

MORETTIN, P. A.; BUSSAB, W. d. O. **Estatística Básica**. São Paulo: Saraiva, 2010. 540 p. ISBN 978-85-02-08177-2.

NIELSON, J.; RALUCA, B. **Mobile Usability**. Berkeley, CA: New Riders, 2013. v. 1. 203 p. ISBN 978-0-321-88448-0.

NORMAN, D. A. **Human-computer interaction**. In: BAECKER, R. M.; BUXTON, W.A. S. (Ed.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987. cap. Some Observations on Mental Models, p. 241–244. ISBN 0-934613-24-9. Disponível em: <<http://dl.acm.org/citation.cfm?id=58076.58097>>.

_____. **The Design of Everyday Things: Revised and Expanded Edition**. Basic Books, 2013. 247 p. ISBN 978-04-65-07299-6.

OECD. **Guidelines on the Protection of Privacy and Transborder Flows of Personal Data**. 2013. Disponível em: <<http://www.oecd.org/sti/ieconomy/oecdguidelinesonthe protectionofprivacyandtransborderflowsofpersonaldata.htm#guidelines>>. Acesso em: 07 Jun. 2018.

_____. **The Organisation for Economic Cooperation and Development**. 2018. Disponível em: <<http://www.oecd.org/about>>. Acesso em: 07 Jun. 2018.

ORENGO, V.; HUYCK, C. A stemming algorithm for the portuguese language. **Proceedings Eighth Symposium on String Processing and Information Retrieval**, p. 186–193, 2001.

PATIL, S.; PAGE, X.; KOBASA, A. With a little help from my friends: Can social navigation inform interpersonal privacy preferences? In: **Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work**. New York, NY, USA:ACM, 2011. (CSCW'11), p.391–394. ISBN 978-1-4503-0556-3. Disponível em: <<http://doi.acm.org/10.1145/1958824.1958885>>

PENTAHO. **Pentaho Faq: What is Data Integration?** 2018. Disponível em: <<http://www.pentaho.com/faq/what-is-data-integration>>. Acesso em: 07 Jun. 2018.

POEPLAU, S. et al. Execute this! analyzing unsafe and malicious dynamic code loading in android applications. In: **Network and Distributed System Security - NDSS**. [S.l.: s.n.], 2014. v. 14, p. 23–26

PONTES, D. de; ZORZO, S. Ppmark: An architecture to generate privacy labels using tf-idf techniques and the rabin karp algorithm. **Advances in Intelligent Systems and Computing**, v. 448, p. 1029–1040, 2016. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84962753128&partnerID=40&md5=4529ac22230c61b90f0d5eaac6feeffd>>.

PORTER, M. F. **The Porter Stemming Algorithm**. 2001. Disponível em: <http://ccl.pku.edu.cn/doubtfire/NLP/Lexical_Analysis/Word_Lemmatization/Porter/Porter%20Stemming%20Algorithm.htm>. Acesso em: 07 Jun. 2018.

_____. An algorithm for suffix stripping. **Program**, MCB UP Ltd, v. 14, n. 3, p. 130–137, 1980.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. [S.l.]: Editora Manole Ltda, 2003.

JIVANI, A. G. et al. A comparative study of stemming algorithms. **International Journal of Computer Technology and Applications - IJCTA**, v. 2, n. 6, p. 1930–1938, 2011.

RODOTÀ, S.; MORAES, M. D. **A vida na sociedade da vigilância: a privacidade hoje**. Rio de Janeiro: Renovar, 2008. 382 p. ISBN 978-85-71-47688-2.

SENGE, P. **A quinta disciplina: arte e prática da organização de aprendizagem**. Nova Cultural, 2002. ISBN 978-85-71-23621-9.

SOUSA, J. S. d. et al. Mulheres digitais: Uma análise da participação das mulheres nos cursos de ciência da computação e engenharia de computação no brasil e na universidade univali. **Anais do Computer on the Beach**, p. 404 – 413, 2017.

TRIOLA, M. F. **Introdução à Estatística**. 12. ed. Rio de Janeiro: LTC, 2017. 836 p. ISBN 978-85-21-63374-7.

WARREN, S. D.; BRANDEIS, L. D. The right to privacy. **Harvard law review**, JSTOR, p. 193–220, 1890. Disponível em: <<http://www.jstor.org/stable/1321160>>.

WESTIN, A. **Privacy and Freedom**. New York, NY, USA: Atheneum, 1967.

WESTIN, A.; SOLOVE, D. **Privacy and Freedom**. Ig Publishing, Incorporated, 2015. 500 p. ISBN 978-19-35-43997-4.

WESTIN, A. F. Privacy in america: An historical and socio-political analysis. **Proceedings of the National Privacy and Public Policy Symposium**, Hartford, CT, USA, 1995.

Apêndice A

TELA DE APRESENTAÇÃO DO TCLE

Figura A.1 – Termo de consentimento livre e esclarecido utilizado na pesquisa.

The image shows a mobile browser interface with a dark blue header and a search bar. The browser address bar shows 'privacidade.info'. The page content is titled 'Bem-Vindo' and contains a 'Termo de Consentimento Livre e Esclarecido' (Informed Consent Form). The form lists 11 numbered points detailing the study's purpose, data collection, confidentiality, and contact information for the researcher, José Santiago Moreira de Mello. At the bottom, there is a blue button labeled 'Eu Concordo' (I Agree).

Termo de Consentimento Livre e Esclarecido

1. Você está sendo convidado para participar de um estudo acadêmico intitulado "Privacidade em Android: confrontando a expectativa dos usuários com o comportamento real dos aplicativos."
2. Você foi selecionado por sua formação acadêmica e sua familiaridade com o uso de aplicativos móveis, no entanto sua participação não é obrigatória.
3. A qualquer momento você poderá desistir de participar do estudo e retirar seu consentimento.
4. Sua recusa em participar do estudo não trará nenhum prejuízo em sua relação com os pesquisadores ou com a instituição onde esse estudo é aplicado.
5. Os objetivos deste estudo estão embasados na possibilidade de promover uma abordagem de privacidade para aplicações móveis levando em conta a percepção dos usuários. Essa percepção deve considerar aspectos sobre qual informação é coletada e o propósito dessa coleta. Essa abordagem será utilizada pelos usuários para apoio na tomada de decisões sobre o uso ou não de determinados aplicativos móveis.
6. Sua participação neste estudo consistirá em responder a um questionário eletrônico, o qual objetiva coletar informações referente a sua expectativa de coleta e uso de dados de seu aparelho celular por parte de algumas aplicações móveis avaliadas nesse estudo.
7. Os benefícios relacionados com a sua participação neste estudo estão ligados a contribuição para a avaliação do trabalho de mestrado proposto e um ganho, tendo esse como sendo um material que poderá ser referenciado no ambiente de pesquisa.
8. Sua participação neste estudo pode envolver algum desconforto relacionado ao tempo despendido com o preenchimento dos questionários apresentados, sendo que faremos o possível para minimizar tais desconfortos. Em relação ao conteúdo dos questionários os mesmos foram planejados de modo a evitar possíveis constrangimentos, e caso ocorram você pode recusar a responder ou mesmo interromper a sua participação a qualquer momento, sem qualquer prejuízo em sua relação com o pesquisador ou com a instituição onde o estudo está sendo conduzido.
9. As informações obtidas através dessa pesquisa serão confidenciais, assegurando que o sigilo sobre sua participação e sobre informações pessoais, como seu nome, serão preservados.
10. Os dados não serão divulgados de forma a possibilitar sua identificação. As informações coletadas não estarão vinculadas à sua identidade.
11. Caso desejar, você poderá receber uma cópia deste termo via e-mail, onde consta o telefone e o endereço de onde encontrar o pesquisador principal, José Santiago Moreira de Mello, podendo tirar suas dúvidas sobre o projeto e sua participação, agora ou a qualquer momento após a participação no estudo.

José Santiago Moreira de Mello
Universidade Federal de São Carlos - Departamento de Computação
Rodovia Washington Luis, km 235, Cep: 13565-905 São Carlos – SP
Tel: (16) 3351-8626
Endereço e telefone do Pesquisador
Rua Padre Rui Cândido da Silva, 1262
19905-152 – Ourinhos / SP
Tel: (11) 96316-3870

Declaro que entendi os objetivos, riscos e benefícios de minha participação na pesquisa e concordo em participar.

Eu Concordo

APÊNDICE B

MODELO FÍSICO DO BANCO DE DADOS

Quadro B.1 - Script de criação do banco de dados.

Modelo Físico: MF_weprivacy.sql

```
1  -- Tabela Aplicativo
2  CREATE TABLE IF NOT EXISTS `weprivacy`.`aplicativo` (
3    `id_aplicativo` INT UNSIGNED NOT NULL AUTO_INCREMENT,
4    `aplicativo` VARCHAR(120) NOT NULL,
5    `empresa` VARCHAR(200) NULL,
6    `processado` TINYINT UNSIGNED NOT NULL DEFAULT 0,
7    `url_politica` VARCHAR(200) NULL,
8    `icone` VARCHAR(200) NULL,
9    `logo` VARCHAR(200) NULL,
10   `descricao_google` VARCHAR(200) NULL,
11   `ranking` TINYINT NULL,
12   PRIMARY KEY (`id_aplicativo`))
13  ENGINE = InnoDB;
14
15  -- Tabela Recurso
16  CREATE TABLE IF NOT EXISTS `weprivacy`.`recurso` (
17    `id_recurso` INT UNSIGNED NOT NULL AUTO_INCREMENT,
18    `recurso` VARCHAR(120) NOT NULL,
19    `icone` VARCHAR(200) NOT NULL,
20    PRIMARY KEY (`id_recurso`))
21  ENGINE = InnoDB;
22
23  -- Tabela Finalidade
24  CREATE TABLE IF NOT EXISTS `weprivacy`.`finalidade` (
25    `id_finalidade` INT UNSIGNED NOT NULL AUTO_INCREMENT,
26    `finalidade` VARCHAR(120) NOT NULL,
27    PRIMARY KEY (`id_finalidade`))
28  ENGINE = InnoDB;
29
30  -- Tabela Pesquisa
31  CREATE TABLE IF NOT EXISTS `weprivacy`.`pesquisa` (
32    `id_pesquisa` INT UNSIGNED NOT NULL AUTO_INCREMENT,
33    `time_inicio` DATETIME NOT NULL,
34    `time_final` DATETIME NOT NULL,
35    PRIMARY KEY (`id_pesquisa`))
36  ENGINE = InnoDB;
```

```
37 -- Tabela Aplicativo_Recurso_Finalidade
38 CREATE TABLE IF NOT EXISTS `weprivacy`.`aplicativo_recurso_finalidade` (
39   `id_app_recurso_finalidade` INT UNSIGNED NOT NULL AUTO_INCREMENT,
40   `id_aplicativo` INT UNSIGNED NOT NULL,
41   `id_recurso` INT UNSIGNED NOT NULL,
42   `id_finalidade` INT UNSIGNED NOT NULL,
43   PRIMARY KEY (`id_app_recurso_finalidade`),
44   INDEX `fk_id_aplicativo_idx` (`id_aplicativo` ASC),
45   INDEX `fk_id_recurso_idx` (`id_recurso` ASC),
46   INDEX `fk_id_finalidade_idx` (`id_finalidade` ASC),
47   CONSTRAINT `fk_tb_aplicativo` FOREIGN KEY (`id_aplicativo`)
48     REFERENCES `weprivacy`.`aplicativo` (`id_aplicativo`)
49     ON DELETE RESTRICT
50     ON UPDATE CASCADE,
51   CONSTRAINT `fk_tb_recurso` FOREIGN KEY (`id_recurso`)
52     REFERENCES `weprivacy`.`recurso` (`id_recurso`)
53     ON DELETE RESTRICT
54     ON UPDATE CASCADE,
55   CONSTRAINT `fk_tb_finalidade` FOREIGN KEY (`id_finalidade`)
56     REFERENCES `weprivacy`.`finalidade` (`id_finalidade`)
57     ON DELETE RESTRICT
58     ON UPDATE CASCADE)
59 ENGINE = InnoDB;
60
61 -- Tabela Aplicativo_Novo
62 CREATE TABLE IF NOT EXISTS `weprivacy`.`aplicativo_novo` (
63   `id_aplicativo_novo` INT UNSIGNED NOT NULL,
64   `aplicativo` VARCHAR(120) NOT NULL,
65   `processado` TINYINT UNSIGNED NOT NULL,
66   PRIMARY KEY (`id_aplicativo_novo`))
67 ENGINE = InnoDB;
68
69 -- Tabela Conforto
70 CREATE TABLE IF NOT EXISTS `weprivacy`.`conforto` (
71   `id_conforto` INT UNSIGNED NOT NULL AUTO_INCREMENT,
72   `descricao` VARCHAR(120) NOT NULL,
73   `indice_likert` TINYINT NOT NULL,
74   `icone` VARCHAR(200) NOT NULL,
75   PRIMARY KEY (`id_conforto`))
76 ENGINE = InnoDB;
77
78 -- Tabela Analise_Proposito
79 CREATE TABLE IF NOT EXISTS `weprivacy`.`analise_proposito` (
80   `id_analise_proposito` INT UNSIGNED NOT NULL AUTO_INCREMENT,
81   `id_aplicativo` INT UNSIGNED NOT NULL,
82   `id_conforto` INT UNSIGNED NOT NULL,
83   `percentual` VARCHAR(6) NOT NULL,
84   `icone` VARCHAR(200) NOT NULL,
85   PRIMARY KEY (`id_analise_proposito`),
86   INDEX `fk_id_aplicativo_idx` (`id_aplicativo` ASC),
87   INDEX `fk_tb_conforto2_idx` (`id_conforto` ASC),
88   CONSTRAINT `fk_tb_aplicativo4` FOREIGN KEY (`id_aplicativo`)
89     REFERENCES `weprivacy`.`aplicativo` (`id_aplicativo`)
90     ON DELETE NO ACTION
91     ON UPDATE NO ACTION,
92   CONSTRAINT `fk_tb_conforto2` FOREIGN KEY (`id_conforto`)
93     REFERENCES `weprivacy`.`conforto` (`id_conforto`)
94     ON DELETE RESTRICT
95     ON UPDATE CASCADE)
96 ENGINE = InnoDB;
```

```
97  -- Tabela Proposito
98  CREATE TABLE IF NOT EXISTS `weprivacy`.`proposito` (
99    `id_pesquisa_proposito` INT UNSIGNED NOT NULL AUTO_INCREMENT,
100   `id_pesquisa` INT UNSIGNED NOT NULL,
101   `id_app_recurso_finalidade` INT UNSIGNED NOT NULL,
102   PRIMARY KEY (`id_pesquisa_proposito`),
103   INDEX `fk_id_app_recurso_finalidade.idx` (`id_app_recurso_finalidade` ASC),
104   INDEX `fk_pesquisa_idx` (`id_pesquisa` ASC),
105   CONSTRAINT `fk_aplicativo_recurso_finalidade` FOREIGN KEY (`id_app_recurso_finalidade`)
106     REFERENCES `weprivacy`.`aplicativo_recurso_finalidade` (`id_app_recurso_finalidade`)
107     ON DELETE NO ACTION
108     ON UPDATE NO ACTION,
109   CONSTRAINT `fk_pesquisa` FOREIGN KEY (`id_pesquisa`)
110     REFERENCES `weprivacy`.`pesquisa` (`id_pesquisa`)
111     ON DELETE NO ACTION
112     ON UPDATE NO ACTION)
113  ENGINE = InnoDB;
114
115  -- Tabela Conforto_Aplicativo
116  CREATE TABLE IF NOT EXISTS `weprivacy`.`conforto_aplicativo` (
117    `id_conforto_aplicativo` INT UNSIGNED NOT NULL AUTO_INCREMENT,
118    `id_pesquisa` INT UNSIGNED NOT NULL,
119    `id_aplicativo` INT UNSIGNED NULL,
120    `id_aplicativo_novo` INT UNSIGNED NULL,
121    `id_conforto` INT UNSIGNED NOT NULL,
122    `grupo` TINYINT UNSIGNED NOT NULL,
123    `time_inicio` DATETIME NOT NULL,
124    `time_final` DATETIME NOT NULL,
125    PRIMARY KEY (`id_conforto_aplicativo`),
126    INDEX `fk_id_aplicativo.idx` (`id_aplicativo` ASC),
127    INDEX `fk_id_conforto.idx` (`id_conforto` ASC),
128    INDEX `fk_id_pesquisa.idx` (`id_pesquisa` ASC),
129    INDEX `fk_id_aplicativo_novo.idx` (`id_aplicativo_novo` ASC),
130    CONSTRAINT `fk_tb_aplicativo3` FOREIGN KEY (`id_aplicativo`)
131      REFERENCES `weprivacy`.`aplicativo` (`id_aplicativo`)
132      ON DELETE NO ACTION
133      ON UPDATE NO ACTION,
134    CONSTRAINT `fk_tb_conforto` FOREIGN KEY (`id_conforto`)
135      REFERENCES `weprivacy`.`conforto` (`id_conforto`)
136      ON DELETE NO ACTION
137      ON UPDATE NO ACTION,
138    CONSTRAINT `fk_tb_pesquisa2` FOREIGN KEY (`id_pesquisa`)
139      REFERENCES `weprivacy`.`pesquisa` (`id_pesquisa`)
140      ON DELETE NO ACTION ON UPDATE NO ACTION,
141    CONSTRAINT `fk_tb_aplicativo_aplicativo_novo` FOREIGN KEY (`id_aplicativo_novo`)
142      REFERENCES `weprivacy`.`aplicativo_novo` (`id_aplicativo_novo`)
143      ON DELETE NO ACTION ON UPDATE NO ACTION)
144  ENGINE = InnoDB;
145
146  -- Tabela Analise_Expectativa
147  CREATE TABLE IF NOT EXISTS `weprivacy`.`analise_expectativa` (
148    `id_analise_expectativa` INT UNSIGNED NOT NULL AUTO_INCREMENT,
149    `id_aplicativo` INT UNSIGNED NOT NULL,
150    `id_recurso` INT UNSIGNED NOT NULL,
151    `percentual` VARCHAR(6) NOT NULL,
152    `icone` VARCHAR(200) NOT NULL,
153    PRIMARY KEY (`id_analise_expectativa`),
154    INDEX `fk_id_recurso.idx` (`id_recurso` ASC),
155    INDEX `fk_tb_aplicativo_idx` (`id_aplicativo` ASC),
156    CONSTRAINT `fk_tb_recurso3` FOREIGN KEY (`id_recurso`)
157      REFERENCES `weprivacy`.`recurso` (`id_recurso`)
158      ON DELETE NO ACTION ON UPDATE NO ACTION,
159    CONSTRAINT `fk_tb_aplicativo5` FOREIGN KEY (`id_aplicativo`)
160      REFERENCES `weprivacy`.`aplicativo` (`id_aplicativo`)
161      ON DELETE NO ACTION ON UPDATE NO ACTION)
162  ENGINE = InnoDB;
```

```
163 -- Tabela Expectativa
164 CREATE TABLE IF NOT EXISTS `weprivacy`.`expectativa` (
165   `id_expectativa` INT UNSIGNED NOT NULL AUTO_INCREMENT,
166   `id_pesquisa` INT UNSIGNED NOT NULL,
167   `id_aplicativo` INT UNSIGNED NULL,
168   `id_aplicativo_novo` INT UNSIGNED NULL,
169   `id_recurso` INT UNSIGNED NOT NULL,
170   `id_finalidade` INT UNSIGNED NOT NULL,
171   PRIMARY KEY (`id_expectativa`),
172   INDEX `fk_id_recurso.idx` (`id_recurso` ASC),
173   INDEX `fk_id_finalidade.idx` (`id_finalidade` ASC),
174   INDEX `fk_id_pesquisa.idx` (`id_pesquisa` ASC),
175   INDEX `fk_id_aplicativo.idx` (`id_aplicativo` ASC),
176   INDEX `fk_id_aplicativo_novo.idx` (`id_aplicativo_novo` ASC),
177   CONSTRAINT `fk_tb_recurso2`
178     FOREIGN KEY (`id_recurso`)
179     REFERENCES `weprivacy`.`recurso` (`id_recurso`)
180     ON DELETE NO ACTION
181     ON UPDATE NO ACTION,
182   CONSTRAINT `fk_tb_finalidade2`
183     FOREIGN KEY (`id_finalidade`)
184     REFERENCES `weprivacy`.`finalidade` (`id_finalidade`)
185     ON DELETE NO ACTION
186     ON UPDATE NO ACTION,
187   CONSTRAINT `fk_tb_pesquisa`
188     FOREIGN KEY (`id_pesquisa`)
189     REFERENCES `weprivacy`.`pesquisa` (`id_pesquisa`)
190     ON DELETE NO ACTION
191     ON UPDATE NO ACTION,
192   CONSTRAINT `fk_tb_aplicativo2`
193     FOREIGN KEY (`id_aplicativo`)
194     REFERENCES `weprivacy`.`aplicativo` (`id_aplicativo`)
195     ON DELETE NO ACTION
196     ON UPDATE NO ACTION,
197   CONSTRAINT `fk_tb_aplicativo_novo`
198     FOREIGN KEY (`id_aplicativo_novo`)
199     REFERENCES `weprivacy`.`aplicativo_novo` (`id_aplicativo_novo`)
200     ON DELETE NO ACTION
201     ON UPDATE NO ACTION)
202 ENGINE = InnoDB;
203
204 -- Tabela Stg_Expectativa
205 CREATE TABLE IF NOT EXISTS `weprivacy`.`stg_expectativa` (
206   `id_stg_expectativa` INT UNSIGNED NOT NULL AUTO_INCREMENT,
207   `id_device` VARCHAR(50) NOT NULL,
208   `datahora_inicio` TIMESTAMP NULL,
209   `datahora_final` TIMESTAMP NULL,
210   `app_name` VARCHAR(120) NULL,
211   `id_recurso` INT NULL,
212   `id_finalidade` INT UNSIGNED NOT NULL,
213   `id_conforto` INT UNSIGNED NOT NULL,
214   PRIMARY KEY (`id_stg_expectativa`),
215   INDEX `fk_stg_expectativa_conforto1_idx` (`id_conforto` ASC),
216   INDEX `fk_stg_expectativa_finalidade1_idx` (`id_finalidade` ASC),
217   CONSTRAINT `fk_stg_expectativa_conforto1`
218     FOREIGN KEY (`id_conforto`)
219     REFERENCES `weprivacy`.`conforto` (`id_conforto`)
220     ON DELETE NO ACTION
221     ON UPDATE NO ACTION,
222   CONSTRAINT `fk_stg_expectativa_finalidade1`
223     FOREIGN KEY (`id_finalidade`)
224     REFERENCES `weprivacy`.`finalidade` (`id_finalidade`)
225     ON DELETE NO ACTION
226     ON UPDATE NO ACTION)
227 ENGINE = InnoDB;
```

```
228 -- Tabela Stg_Proposito
229 CREATE TABLE IF NOT EXISTS `weprivacy`.`stg_proposito` (
230   `id_stg_proposito` INT UNSIGNED NOT NULL AUTO_INCREMENT,
231   `id_device` VARCHAR(50) NOT NULL,
232   `datahora_inicio` TIMESTAMP NULL,
233   `datahora_final` TIMESTAMP NULL,
234   `id_aplicativo` INT UNSIGNED NOT NULL,
235   `id_conforto` INT UNSIGNED NOT NULL,
236   PRIMARY KEY (`id_stg_proposito`),
237   INDEX `fk_stg_proposito_aplicativo1_idx` (`id_aplicativo` ASC),
238   INDEX `fk_stg_proposito_conforto1_idx` (`id_conforto` ASC),
239   CONSTRAINT `fk_stg_proposito_aplicativo1`
240     FOREIGN KEY (`id_aplicativo`)
241     REFERENCES `weprivacy`.`aplicativo` (`id_aplicativo`)
242     ON DELETE NO ACTION
243     ON UPDATE NO ACTION,
244   CONSTRAINT `fk_stg_proposito_conforto1`
245     FOREIGN KEY (`id_conforto`)
246     REFERENCES `weprivacy`.`conforto` (`id_conforto`)
247     ON DELETE NO ACTION
248     ON UPDATE NO ACTION)
249 ENGINE = InnoDB;
```
