

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL MESTRADO

RODRIGO SILVA PEREIRA

*CROWDHEALTH: UM SISTEMA DE RECOMENDAÇÃO DE CLÍNICAS DE SAÚDE
NUM CONTEXTO SMART-HEALTH USANDO CROWDSOURCING*

SÃO LEOPOLDO
2016

Rodrigo Silva Pereira

*CROWDHEALTH: UM SISTEMA DE RECOMENDAÇÃO DE CLÍNICAS DE SAÚDE
NUM CONTEXTO SMART-HEALTH USANDO CROWDSOURCING*

Dissertação apresentada como requisito parcial
para a obtenção do título de Mestre pelo
Programa de Pós-Graduação em Computação
Aplicada da Universidade do Vale do Rio dos
Sinos — UNISINOS

Orientador:
Prof. Dr. Rodrigo da Rosa Righi

São Leopoldo
2016

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Pereira, Rodrigo Silva

CrowdHealth: Um sistema de Recomendação de Clínicas de Saúde num contexto *Smart-Health* usando *crowdsourcing* / Rodrigo Silva Pereira — 2016.

57 f.: il.; 30 cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, 2016.

“Orientador: Prof. Dr. Rodrigo da Rosa Righi, Unidade Acadêmica de Pesquisa e Pós-Graduação”.

1. Sistemas de Recomendação. 2. Crowdsourcing. 3. Saúde Inteligente. 4. Registros de Saúde Pessoais. I. Título.

CDU 004.732

Bibliotecária responsável: Fulana da Silva — CRB 12/3456

AGRADECIMENTOS

Agradeço a minha família, especialmente a mãe Flávia, tia Fátima e minha vó Zilda. Pilares que fundamentaram o que sou hoje, por sempre acreditarem em mim e me apoiarem nas minhas decisões.

Ao meu orientador e xará Prof. Rodrigo Righi, e todo corpo docente do PIPCA, pela dedicação prestada durante todos estes meses.

Aos colegas da turma de 2014/1, Natália Navarro, Willian Hart, Vinicius Facco, Marcelo Telles, Paulo Henrique Santini, Joel Gonsalves, Milene Martini, Gustavo Rostirolla, Felipe Plets, pelo vínculo de amizade, cumplicidade e grupo de apoio que se criou.

A DATACOM e meus ex-colegas de trabalho pelo apoio e incentivo.

Enquanto escrevo, Manuela, minha filha de dois anos e pouco, lambuza a tela do meu notebook que recebe, mudo, meus falares, com sua mãozinha suja de pipoca. Volto rapidamente das lembranças passadas para a alegria de ter terminado este trabalho e, enfim, poder dedicar mais de mim a ela.

Temos um código de comunicação próprio, e quando ela acha que estou perdendo muito tempo concentrado numa atividade, vem pro meu lado e cria uma situação desafiadora, como lambuzar a tela onde escrevo e cair na gargalhada com o feito. Desligo tudo. Sigo seus passos, seu verdadeiro e irresistível sorriso, este sim, bálsamo milagroso e infalível contra os efeitos do tempo e da constante busca pela lucidez, sonho cada vez mais difícil.

RESUMO

Com a emergência do *crowdsourcing* junto a difusão mundial de *smartphones* esforços recentes e pesquisas importantes sobre o uso de *crowdsourcing* na área da saúde ou ainda *smart-health* visam auxiliar na melhoria hábitos de saúde, construção de históricos médicos pessoais de longo prazo, análise e revisão de dados médica, controle de dietas alimentares, gerenciamento do estresse, análise e comparação de informações e assistência em tempo real para catástrofes. Porém, nenhum deles usou de *crowdsourcing* para recomendação de centros clínicos de saúde. Segundo Chatzimilioudis *crowdsourcing* refere-se "a um modelo distribuído de solução de problemas em que uma multidão de tamanho indefinido é contratada para resolver um problema complexo através de um convite aberto". Neste âmbito, este trabalho apresenta um modelo de sistema de recomendação de centros clínicos de saúde, chamado *CrowdHealth*. A principal contribuição do modelo de sistema de recomendação de centros clínicos é possibilitar a criação de uma relação ganha-ganha entre seus usuários que podem ser cidadãos, médicos ou ainda entidades ligadas ao governo. Na literatura encontramos alguns trabalhos que carecem a abordagem do uso de *crowdsourcing* como fonte de dados para recomendação de centros clínicos de saúde. Nós desenvolvemos um protótipo de aplicação baseada no modelo de sistema de recomendação de centros clínicos de saúde para proporcionar uma visão do que seria uma aplicação baseada no modelo de sistema de recomendação de centros clínicos de saúde. Para avaliar o nosso modelo, apresentamos um cenário hipotético baseado numa possível aplicação para mensurar a percepção dos usuários quanto a utilidade dos centros clínicos de saúde. Os cenários descritos levavam em consideração os seguintes critérios: (1) a distância entre do usuário ao centro clínico, (2) a avaliação dos usuários em relação ao atendimento recebido nos centros clínicos e (3) o tempo de atendimento informado pelos usuários. Desta forma realizamos uma simulação de requisições de recomendações de usuários usando um *dataset* real contendo informações do *Foursquare*. O arquivo do *dataset* possuía 227428 *check-in's* na cidade de Nova Iorque, EUA. O arquivo, foi dividido em duas partes, onde a primeira representava os *check-in's* realizados pelos usuários nos centros clínicos, e a segunda representava usuários requisitando por recomendações de centros clínicos em outros locais. Assim, foram criadas funções para simular os processos de cálculo do tempo de atendimento e avaliação dos centros clínicos por parte dos usuários. Também simulou-se usuários requisitando por recomendações de centros clínicos em outros locais. Então, medimos precisão e recuperação dos centros clínicos de saúde sugeridos para cada usuário. Obtivemos valores médios de 57,5% e 61,33% para precisão e recuperação, respectivamente. Com isso, nossa avaliação retrata que centros clínicos de saúde recomendados por uma aplicação baseada no *CrowdHealth* poderiam aumentar beneficemente a utilidade de centros clínicos de saúde recomendados para os usuários.

Palavras-chave: Sistemas de Recomendação. Crowdsourcing. Saúde Inteligente. Registros de Saúde Pessoais.

ABSTRACT

With the emergence of crowdsourcing with the worldwide spread of smartphones recent efforts and important research on the use of crowdsourcing in health or smart-health are intended to assist in improving health habits, construction of historical long-term medical personnel, medical analysis and data review, control diets, stress management, analysis and comparison of information and real-time assistance for disasters. However, none of them used the crowdsourcing for recommendation clinical health centers. In this context, this paper presents a model of clinical health centers recommendation system called CrowdHealth. The main contribution of clinical health centers recommendation system model is possible to create a win-win relationship between its users that can be citizens, doctors or entities linked to the government. In the literature we find some papers that require the use of crowdsourcing as a data source for recommendation clinical health centers approach. We have developed a prototype application based on clinical health centers recommendation system model to provide a vision of what would be an application based on the clinical health centers recommendation system model. To evaluate our model, we present a hypothetical scenario based on a possible application to measure the perception of users and the utility of clinical health centers. The scenarios described took into consideration the following criteria: (1) the distance from the user to the clinical center, (2) the evaluation of other users on the service received in the clinical centers and (3) the time of service reported by users. Thus we performed a simulation of user requests recommendations using a real dataset containing information of Foursquare. The file dataset had 227428 check in's in New York City, USA. The file was divided into two parts, where the first represented the textit check in 's performed by users in clinical centers, and the second represented by requesting users polyclinics recommendations elsewhere. Thus, functions were created to simulate service time calculation and evaluation processes of polyclinics by users. Also users was simulated by ordering polyclinics recommendations elsewhere. So we measure precision and recall of health clinical centers suggested for each user. Average values obtained from 57.5 % and 61.33 % for precision and recall, respectively. Thus, our assessment that portrays clinical health centers recommended by an application based on CrowdHealth could increase beneficially the usefulness of clinical health centers recommended for users.

Keywords: *Recommendations Systems, Crowdsourcing, Smart-Health, Personal Health Record.*

LISTA DE FIGURAS

Figura 1:	<i>Crowdsourcing</i> com <i>smartphones</i>	16
Figura 2:	Fenômeno da Cauda Longa	18
Figura 3:	Taxonomia de Sistemas de Recomendação	19
Figura 4:	Diagrama de conjunto entre os planos da saúde e Smart City.	21
Figura 5:	Modelo do sistema Demorô	23
Figura 6:	Telas do sistema <i>HealthGem</i>	24
Figura 7:	Telas do sistema <i>HealthGem</i>	25
Figura 8:	Arquitetura do sistema PHR de Song	26
Figura 9:	Arquitetura do sistema de Papageorgiou.	28
Figura 10:	Arquitetura do sistema <i>Fihuo</i>	30
Figura 11:	Evolução dos Modelos de cuidados de saúde.	32
Figura 12:	Modelo do <i>CrowdHealth</i>	35
Figura 13:	Comunicação do <i>CrowdHealth</i> com os usuários	36
Figura 14:	Arquitetura do <i>CrowdHealth</i>	39
Figura 15:	Protótipo da Aplicação do Cliente	40
Figura 16:	Protótipo da Aplicação do Cliente	41
Figura 17:	Comparativo entre Resultados - Precision	50
Figura 18:	Comparativo entre Resultados - Recall	50

LISTA DE TABELAS

Tabela 1:	Comparação entre trabalhos correlatos	33
Tabela 2:	Serviços oferecidos pelo componente <i>Web Service</i>	42
Tabela 3:	Métricas de Avaliação	44
Tabela 4:	Resultados Absolutos	49
Tabela 5:	<i>CrowdHealth</i> vs Trabalhos Relacionados	52

LISTA DE SIGLAS

API	Interface de Processamento da Aplicação, do inglês <i>Application Process Interface</i> .
CPU	Unidade Central de Processamento, do inglês <i>Central Processing Unit</i> .
CTR	Taxa através de cliques, do inglês <i>Click-Through Rate</i> .
DMRQ	Desvio Médio da Raiz Quadrada (ou RMSD em inglês).
EHR	Registro de saúde eletrônico, do inglês <i>Electronic health record</i> .
EQM	Erro Quadrático Médio (EQM, ou MSE em inglês).
JSON	Notação de Objetos <i>JavaScript</i> do inglês <i>JavaScript Object Notation</i> .
Km	Kilometros.
PHR	Registros de Saúde Pessoal, do inglês <i>Personal Health Record</i> .
TIC	Tecnologia da Informação e Comunicação
XML	<i>eXtensible Markup Language</i> ou, em português Linguagem Extensível de Marcação.
UTC	Tempo Universal Coordenado, do inglês <i>Universal Time Coordinated</i> .

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Motivação	11
1.2	Questão de Pesquisa e Objetivos	13
1.3	Organização do Texto	13
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Sistemas <i>Crowdware</i>	14
2.2	Crowdsourcing	14
2.3	<i>Crowdsourcing com smartphones</i>	16
2.4	Sistemas EHR e PHR	17
2.5	O Fenômeno da Cauda Longa	18
2.6	Sistemas de Recomendação	19
2.7	<i>Smart Cities</i>	20
2.8	<i>Smart Health</i>	20
2.9	Considerações sobre o Capítulo	21
3	TRABALHOS RELACIONADOS	22
3.1	Metodologia de Pesquisa	22
3.2	Leveraging The Crowd Collaboration To Monitor The Waiting Time Of Day-To-Day Services	22
3.3	The Use Of Social Networks In Aid Of Personal Health Monitoring	23
3.4	Empowering Patients Using Cloud Based Personal Health Record System	25
3.5	CROWDHELP: M-Health App For Emergency Response Improvement Through Crowdsourced And Sensor-Detected Information	26
3.6	PEOPLESAVE: Recommending Effective Drugs Through Web Crowdsourcing	27
3.7	SMART FOOD: Crowdsourcing Of Experts In Nutrition And Non-Experts In Identifying Calories Of Meals Using Smartphone As A Potential Tool Contributing To Obesity Prevention And Management	27
3.8	Personalising And Crowdsourcing Stress Management In Urban Environments Via S-Health	28
3.9	FIHUO: A Moblie S-Health Service Platform	29
3.10	Comparação entre os Trabalhos Relacionados	30
4	<i>CROWDHEALTH</i>	34
4.1	Premissas	34
4.2	Estratégias e Decisões de Projeto	36
4.3	Formalização do Problema	37
4.4	Arquitetura	38
4.4.1	Camada do Cliente	39
4.4.2	Camada de API	40
4.4.3	Camada de Aplicação	40
4.4.4	Camada de Banco de Dados	41
4.4.5	Camada de <i>Data Mining</i>	41

5	METODOLOGIA DE AVALIAÇÃO E RESULTADOS	43
5.1	Decisões de Projeto	43
5.2	CrITÉrios de AvaliaÇo	43
5.3	ImplementaÇo	45
5.4	Pesquisa as Cegas	48
5.5	Arquitetura dos Experimentos	48
5.6	Resultados	49
6	CONSIDERAÇES FINAIS	51
6.1	Concluses	51
6.2	Trabalhos Futuros	51
6.3	ContribuiÇes	53
	REFERNCIAS	54

1 INTRODUÇÃO

Recentemente, Jeff Howe introduziu o conceito de *crowdsourcing* que pode basicamente ser entendido por *uma fonte de informações oriundas de uma multidão*. Segundo Chatzimilioudis, *crowdsourcing* refere-se a "um modelo distribuído de solução de problemas em que uma multidão de tamanho indefinido é contratada para resolver um problema complexo através de um convite aberto" (CHATZIMILIOUDIS et al., 2012). A estratégia de alavancar multidões para solucionar problemas difíceis ou não triviais tem sido investigada nos últimos anos através de diversos pesquisadores.

Também recentemente Solanas, introduziu um outro conceito, o de *Smart Health* (*s-health* ou saúde inteligente em Português), no qual recursos de uma cidade inteligente são explorados para facilitar a previsão de serviços de saúde para os cidadãos (SOLANAS et al., 2014). O termo *cidades inteligentes* vem do inglês, *Smart Cities*. Surgiu através das transformações que estão acontecendo nas cidades com propósito de torna-las mais sustentáveis e aumentar a eficiência de recursos que nela existem (ELMANGOUSH et al., 2013). Esta transformação é altamente dependente da Tecnologia da Informação e Comunicação (TIC), onde as informações de uma grande rede interconectada de sensores e outros dispositivos exibem diversos aspectos da vida urbana em tempo real (NAPHADE et al., 2011).

Ainda na área da saúde, *Personal Health Records* (ou Registros de Saúde Pessoal em Português), estão se tornando um campo crescente e a principal plataforma, com fornecedores de dispositivos que oferecem soluções alternativas que abordam tudo, desde softwares de acompanhamento manual até dispositivos baseados em sensores de rastreamento automatizado (ANTHONY; PHILLIP; MAE, 2015). Embora se tenha vários sistemas de monitoramento e gestão de saúde atualmente, o uso de PHR's está limitado a apenas uma parte relativamente pequena de usuários. Uma das principais razões, segundo Anthony, é que as pessoas veem a saúde, a integração e atualização de informações sobre médicos, clínicas e hospitais, como algo necessário, apenas durante os períodos de doença, em vez de uma responsabilidade que tenha que ser monitorada e mantida pessoalmente e continuamente. Isso leva os usuários a não atualizar suas informações de saúde nos PHR/s (ANTHONY; PHILLIP; MAE, 2015). Outra questão levantada por Cerrato, é que as pessoas são preguiçosas na tomada de medidas preventivas de sua saúde. Eles veem na saúde algo como "ter alguém para torná-lo melhor, e não algo sobre uma responsabilidade pessoal"(CERRATO, 2012).

1.1 Motivação

Com a emergência de *crowdsourcing* junto a difusão mundial de *smartphones*, esforços recentes e pesquisas importantes sobre o uso de *crowdsourcing* na área da saúde ou ainda em *smart-health* visam auxiliar seus usuários em diversos aspectos. *Smartphones* podem revelar-se potenciais fontes de *crowdsourcing* e permitir que os usuários contribuam de forma transparente

para resolução de problemas novos e complexos. Atualmente, existem nas grandes cidades diversos centros clínicos de saúde e há também diversas informações computacionais disponíveis através de *smartphones* que poderiam ser utilizadas para auxiliar as pessoas em um momento de necessidade de atendimento médico. Cidades inteligentes permitem que esses dados sejam usados para fornecer uma recomendação que pode ser mais útil do que se a mesma pessoa escolhesse as cegas um centro clínico de saúde qualquer.

O trabalho de Papageorgiou, explora os potenciais de uma cidade inteligente utilizando dados oriundos de fontes *crowdsourcing* para resolver questões de saúde pública, entretanto o objetivo do modelo apresentado no trabalho é o gerenciamento do estresse (PAPAGEORGIOU; ZIGOMITROS; PATSAKIS, 2015). Já Anthony explora o uso de *crowdsourcing* na saúde em seu trabalho através do *HealthGem* com o objetivo de monitorar e manter os usuários motivados em levar um estilo de vida saudável (ANTHONY; PHILLIP; MAE, 2015).

Poulymenopoulou cita que, um dos principais desafios enfrentados durante a concepção e implementação do sistema de saúde apresentado em seu trabalho, foi a questão da motivação dos participantes para se juntarem ao esforço da análise de saúde e dar permissões para extrair resumos de pacientes, a partir de dados armazenados localmente para serem transferidos para a nuvem (POULYMENOPOULOU et al., 2015). Oliveira propôs uma aplicação *mobile* para monitoramento de tempo de espera de serviços oferecidos numa cidade de forma genérica. Porém, não é citado como as soluções poderiam motivar os usuários a compartilhar seus dados com as demais entidades de seu sistema. A proposta do *CrowdHealth* é semelhante ao trabalho de Oliveira, porém especificamente na área da saúde (OLIVEIRA et al., 2015).

No artigo de Majethia, é apresentado um sistema de análise de revisão médica *crowdsourcing* para a recomendação e interrupção do uso de drogas com base na análise contextual dos comentários das multidões na Internet, chamado *PeopleSave* (MAJETHIA, 2016). O estudo de Moorhead, mostra um sistema para *smartphones* para controle e gerenciamento de dietas alimentares (MOORHEAD; BOND, 2014).

Diferente desses trabalhos, o objetivo do *CrowdHealth* é auxiliar na tomada de decisão dos usuários acerca de qual centro clínico de saúde é mais aconselhável que ele vá no momento em que ele precisa. Não é possível apresentar todos os centros clínicos de saúde disponíveis para o usuário. Nem podemos esperar que os usuários conheçam todos os centros clínicos de saúde que podem preferir (RAJARAMAN; ULLMAN, 2011). Assim, selecionar quais dos centros clínicos de saúde disponíveis na cidade é o mais adequado para cada usuário é uma tarefa complexa.

Através da análise do tempo de espera (e outras variáveis) informado pelos usuários pode-se elaborar um modelo que permite criar uma relação ganha-ganha entre seus usuários. Ao permitir o compartilhamento de suas informações os usuários ganham tomando conhecimento das informações fornecidas por outros.

1.2 Questão de Pesquisa e Objetivos

A área da saúde ainda foi pouco explorada no âmbito das *crowdsourcing*, ou ainda, existem lacunas de pesquisa que podem ser preenchidas a partir deste trabalho. A ideia central aqui é apresentar uma solução, que seria melhor aproveitada se integrada a um PHR, usando fontes de dados *crowdsourcing* num ambiente de *smarth-health*. Portanto, é nesse contexto de *smarth-health* e *crowdsourcing* que esta dissertação está inserida, tendo como objetivo principal responder a seguinte questão:

Como seria uma solução *smarth-health* que motivasse multidões a utilizar e compartilhar as informações de seus PHR's, a fim de contribuir com a saúde pública através de uma abordagem de *crowdsourcing*?

Tal questão pode ser desdobrada nos seguintes objetivos específicos:

- Analisar como os demais pesquisadores tratam dos trabalhos semelhantes a este.
- Selecionar os trabalhos mais relevantes e encontrar lacunas de pesquisa nestes trabalhos.
- Apresentar um modelo científico de solução que motive multidões a compartilhar seus dados de saúde pessoais por meio de uma relação ganha-ganha entre os usuários envolvidos.
- Extrair conhecimentos das informações obtidas através do compartilhamento dos dados de saúde dos usuários.

1.3 Organização do Texto

Este trabalho está organizado da seguinte forma: No próximo capítulo é apresentado a fundamentação teórica necessária para entendimento desta dissertação. No capítulo 3 são mostrados os trabalhos relacionados, como eles foram escolhidos e uma análise de como os principais trabalhos associados ao tema de pesquisa abordam o uso de *crowdsourcing* na saúde e ainda as soluções *s-health* existentes. Ao final do capítulo, os trabalhos são confrontados em uma tabela onde é demonstrado a lacuna de pesquisa encontrada. O capítulo 4 apresenta o modelo do *CrowdHealth*, juntamente com sua arquitetura, seus objetivos e as estratégias utilizadas na criação do modelo. O capítulo mostra como foi feita a implementação, os experimentos e a metodologia de avaliação. Por fim, o capítulo 5 contém as conclusões desta dissertação, as contribuições científicas e também algumas questões que futuramente poderão ser tema de pesquisa para outros pesquisadores.

2 FUNDAMENTAÇÃO TEÓRICA

Serão apresentados neste capítulo alguns conceitos importantes para entendimento desta dissertação. O primeiro deles, sistemas *crowdware*, serve de introdução para o conceito de são sistemas *crowdsourcing*. Em seguida são apresentados os tipos de *crowdsourcing* existentes e sua origem. Também é exibido nessa mesma seção o conceito de *crowdsourcing* através de *smartphones* para entendimento de soluções *mobile health* (saúde móvel, em Português). Mais adiante, na seção 2.4, é apresentado o conceito de sistemas na área de saúde tais como EHR (*Electronic Health Record*) e PHR (*Personal Health Record*). E por fim, mas não menos importante, na seção 2.6 é apresentado o conceito e os principais tipos de sistemas de recomendação existentes.

2.1 Sistemas *Crowdware*

O termo *crowdware* foi inspirado no termo *groupware*, o qual sugere princípios e ideias de comunicação e colaboração dentro de um contexto compartilhado que ajude um grupo a realizar uma tarefa em conjunto (ZHANG et al., 2012). Sistemas *crowdware* introduzem novos contextos onde multidões (*crowds*) podem comunicar, colaborar e interagir em uma variedade de atividades e tarefas ligadas à *Internet* de forma consciente. Sistemas *crowdware* podem ser descritos como uma classe de sistema de auxílio real e virtual de multidões, herdando os principais componentes de *groupware*, através da *Web 2.0* (PRILLA; RITTERSKAMP, 2008) e *cloud computing* (computação em nuvens, em Português) (AGRAWAL; DAS; El Abbadi, 2011), para prover serviços avançados em qualquer lugar, a qualquer hora, conectando pessoas em domínios heterogêneos (SCHNEIDER et al., 2012). *Blogs*, *wikis*, serviços de redes sociais, sistemas colaborativos de multidão baseados na *Web* e "artefatos inteligentes"(dispositivos baseados no computador não tradicionais que permitem trabalho remoto, ex: *tablets*, *smartphones*, sistemas de monitoramento de saúde como *wearables*) são todas formas de *crowdware* (SCHNEIDER et al., 2012).

2.2 Crowdsourcing

O termo *crowdsourcing* está diretamente relacionado a questões discutidas neste trabalho. Ele apareceu pela primeira vez num artigo escrito pelo jornalista Jeff Howe, editor da revista americana *Wired* em 2006. Howe foi o primeiro a apontar a tendência. O termo *crowdsourcing* também é definido por *Merriam-Webster*¹ como: "*Ato de reunir habilidades perceptivas, cognitivas ou atuantes de muitas pessoas para alcançar um resultado bem definido, como resolver um problema, classificando um conjunto de dados, ou produzir uma decisão*". A tradução literal de *Crowdsourcing* vem da união de duas palavras do inglês, *crowd*, que significa multidão e

¹*Merriam-Webster* é uma empresa americana que publica livros de consulta, especialmente dicionários.

source, que significa fonte, origem, manancial, raiz e também fonte de informações. Portanto, *crowdsourcing* poderia ser traduzido para algo como "*fonte de informações oriundas de uma multidão*".

Já para Rouse, *crowdsourcing* pode ser definido como "*a prática de terceirização de tarefas para um grupo externo amplo, vagamente definido de pessoas*" (ROUSE, 2016). *Crowdsourcing* é um conceito aberto, ainda em definição. Entretanto, para este trabalho será utilizado a definição de Chatzimilioudis de que *crowdsourcing* refere-se "*a um modelo distribuído de solução de problemas em que uma multidão de tamanho indefinido é contratada para resolver um problema complexo através de um convite aberto*" (CHATZIMILIOUDIS et al., 2012).

Brabham em seu livro *Crowdsourcing* (BRABHAM, 2013) criou a seguinte tipologia baseada na solução de problemas através de uma abordagem *crowdsourcing*:

- **Descoberta de Conhecimento e Gestão** - para problemas de gestão da informação, onde uma organização mobiliza uma multidão para encontrar e reunir informações. Ideal para a criação de recursos coletivos.
- **Inteligência Humana Distribuída em Multitarefa** - para questões de gestão da informação, onde uma organização tem um conjunto de informações em mãos e mobiliza uma multidão para processar ou analisar a informação. É usado no processamento de grandes conjuntos de dados que os computadores não podem facilmente fazer.
- **Pesquisa Broadcast** - para problemas relacionados a definição de ideias, onde uma organização mobiliza uma multidão para chegar a uma solução de um problema que tem uma resposta correta e objetiva. Ideal para a resolução de problemas científicos.
- **Ponto-vetados de Produção Criativa** - também está relacionado a definição de problemas de ideias, porém aqui as multidões são mobilizadas para chegar a solução de uma questão que tenha uma resposta subjetiva ou dependente do apoio público. Ideal para projetos estéticos ou de mapeamento de problemas políticos.

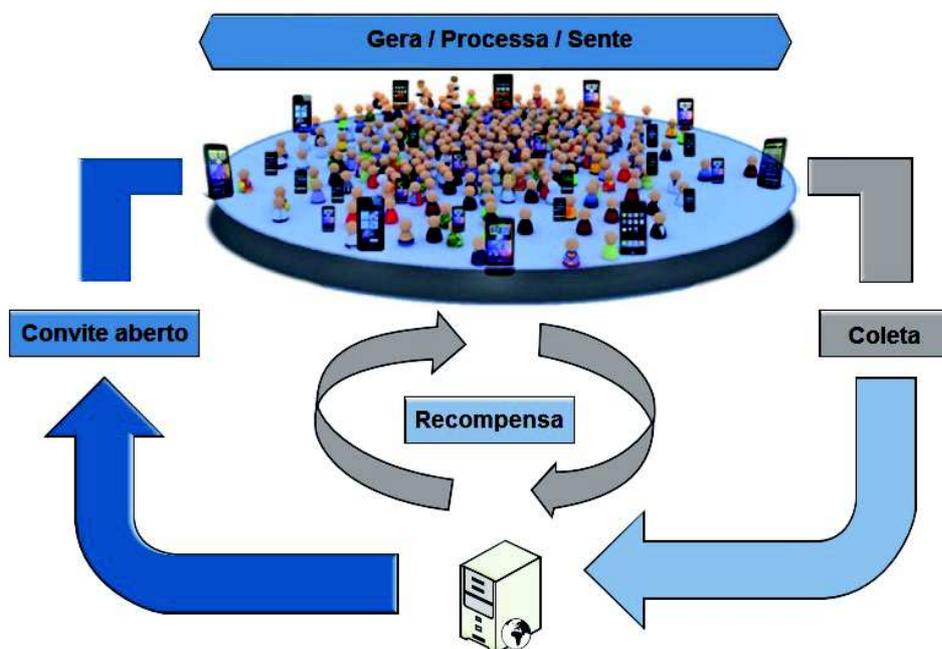
Como parte do seu trabalho, o estudo de Quinn e Bederson demonstra que o uso de *crowd* é incentivado através de pagamento, altruísmo, engajamento, reputação (QUINN; BEDERSON, 2011). Além da tipologia de Brabham, para fins de exemplo é possível citar publicações relacionadas a *crowd computing* nas áreas de planejamento estratégico (MEDEIROS et al., 2012), categorização de requisitos (NASCIMENTO et al., 2012), *crowdsourcing* na música (GOMES et al., 2013a) (GOMES et al., 2012) (GOMES et al., 2013b), ciência do cidadão (ANTELIO et al., 2012) (MORAES et al., 2014) e jornalismo movido pela multidão (SCHNEIDER; De Souza; LUCAS, 2014). Além disso, os recentes avanços na tecnologia de telefonia móvel estimularam o nascimento de um novo paradigma para realização de detecção em grande escala conhecido na literatura como *sensoriamento participativo* (do inglês *participatory sensing*) (CHEN; WANG, 2013), (MORAES et al., 2014).

2.3 Crowdsourcing com smartphones

Smartphones podem revelar-se potenciais fontes de *crowdsourcing* e permitir que os usuários contribuam de forma transparente para resolução de problemas novos e complexos (CHATZIMILIOUDIS et al., 2012). Embora o *crowdsourcing* ainda não tenha penetrado totalmente na força de trabalho móvel, o uso generalizado de *smartphones* - que estão sempre conectados - acabará por revelar todo o potencial desta nova abordagem de resolução de problemas. *Smartphones* oferecem uma ótima plataforma para estender aplicações de *crowdsourcing* existentes baseadas na *Web* para uma multidão ampliando a contribuição, fazendo com que ela seja mais fácil e onipresente. Além disso, as capacidades multissensoriais dos *smartphones* - incluindo os sensores de geolocalização, luz, movimento, visual e auditivo - oferecem uma variedade de maneiras novas e eficientes para coleta oportunista de dados, permitindo novas aplicações de *crowdsourcing* (CHATZIMILIOUDIS et al., 2012).

A Figura 1 ilustra um sistema *crowdsourcing* através de *smartphones* que são utilizados pelos usuários para coletar, processar e/ou sentir informações relevantes para solução de um problema, o qual eles estão engajados em resolver através de um convite aberto. Essas informações são coletadas e enviadas para *web* e isso gera algum tipo de recompensa que estimula os usuários a continuar utilizando o sistema.

Figura 1: *Crowdsourcing* com *smartphones*.



Fonte: Adaptada de (CHATZIMILIOUDIS et al., 2012)

2.4 Sistemas EHR e PHR

Um Registro de Saúde Eletrônico (do Inglês, *Electronic Health Record*) é "uma coleção longitudinal de informações eletrônicas de saúde a respeito de pacientes individuais e populações" (GUNTER; TERRY, 2005). O principal objetivo de sistemas EHR é prover informações clínicas a respeito do paciente onde suas necessidades para consultar, independentemente do tipo de localização de origem, ajudam a evitar erros clínicos ou esforços e recursos duplicados. Portanto, trata-se supostamente de um mecanismo para integrar informações e cuidados de saúde com o propósito de melhorar a qualidade de saúde (OFFICE, 2008).

Conforme definido pelo Coordenador de Público de Tecnologia e Informação de Saúde dos EUA, o uso significativo de EHR's podem beneficiar não somente instituições médicas mas também o público em geral de diversas formas, dentre elas, a transição de um modelo centrado na clínica para um modelo centrado no paciente. Graças a interoperabilidade dos sistemas EHR - que são compatíveis com o critérios de uso significativo, tomar controle dos nossos próprios registros não é mais uma visão otimista. Contudo, o acesso dos usuários ao seus EHR tem benefícios quando não se tem uma maneira de entregar as condições de saúde durante as visitas médicas (SONG; HONG; PAK, 2015).

Ainda segundo Song, nos EUA, cerca de 100,000 pacientes morrem todos os anos e cerca de mais de 1,5 milhão são afetados por erros médicos. Este grande número pode ser reduzido significativamente através do engajamento de tecnologias da informação no compartilhamento de dados médicos (como por ex.: históricos médicos pessoais, exames, tratamentos médicos) (SONG; HONG; PAK, 2015). Por esta razão, tem crescido o interesse no desenvolvimento de sistemas PHRs apropriados conforme mostrado nas publicações (LEE et al., 2013), (DO et al., 2011), (ADIDA et al., 2010).

Segundo Song, um sistema EHR possui duas partes principais: um EHR e um PHR (*Personal Health Record*, ou Registro de Saúde Pessoal em Português) (SONG; HONG; PAK, 2015).

- **EHR:** Uma informação relacionada a um registro eletrônico de um indivíduo que possui padrões de interoperabilidade reconhecidos nacionalmente e que possam ser criados, gerenciados e consultados por médicos e agentes autorizados através de mais de uma organização de cuidados de saúde.
- **PHR:** São aplicações eletrônicas através das quais indivíduos podem manter e gerenciar suas informações de saúde (juntamente com outros por quem eles são autorizados) em um ambiente privado, seguro e confidencial.

Sawyer cita em seu estudo que o objetivo do PHR é para encorajar os pacientes a tornarem-se mais envolvidos e informados como verdadeiros parceiros em seus cuidados, fazendo escolhas positivas para melhorar ou manter sua saúde. Além disso, devido ao aumento da prevalência de condições de longo prazo, o envolvimento dos pacientes em seus cuidados é visto como

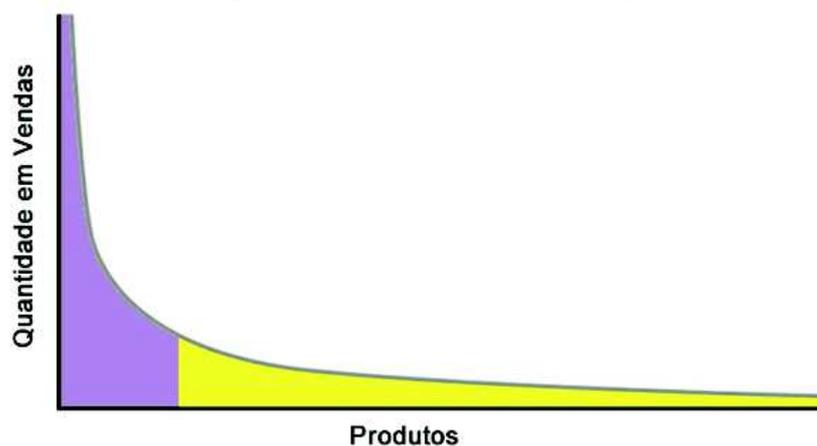
economia potencial de custos, e como tal os PHR's tornaram-se uma prioridade estratégica (SAWYER; AYERS; FIELD, 2010).

2.5 O Fenômeno da Cauda Longa

Rajaraman apresenta, em seu livro *Mining of Massive Datasets*², o Fenômeno da Cauda Longa, que faz a recomendação de sistemas necessária. Sistemas de entregas físicas são caracterizados pela escassez de recursos. Lojas de materiais de construção possuem espaços limitados nas prateleiras, e podem mostrar aos clientes somente uma pequena fração de todas as escolhas possíveis. Por outro lado, lojas *online* podem exibir qualquer coisa que exista a disposição dos clientes. Assim, uma livraria física pode ter vários milhares de livros em suas prateleiras, porém a *Amazon*³ possui milhões de livros a venda. Um jornal físico pode imprimir várias dezenas de artigos por dia, enquanto serviços *online* oferecem milhares de notícias por dia (RAJARAMAN; ULLMAN, 2011).

Recomendações no ambiente físico são bastante simples. Primeiramente, não é possível adequar o armazenamento para cada cliente individualmente. Assim, a escolha do que é disponibilizado é regulada apenas pelos números acumulados. Geralmente, uma livraria mostrará somente os livros que são de maior interesse popular, e uma revista publicará somente artigos que acredita-se que a maioria das pessoas terão interesse em ler. No primeiro caso, o número de vendas é que governa as escolhas, no segundo, o julgamento editorial.

Figura 2: Fenômeno da Cauda Longa



Fonte: Adaptado de (RAJARAMAN; ULLMAN, 2011)

Essa distinção entre os ambientes físicos e o *online* tem sido chamado de Fenômeno da Cauda Longa, e é apresentado na Figura 2. O eixo vertical representa a popularidade (o número de vezes que um item é escolhido). Os artigos são ordenados sobre o eixo horizontal de acordo

²Mineração de conjuntos massivos de dados, em português

³http:

www.amazon.com

com a sua popularidade. Instituições físicas fornecem apenas os itens mais populares para a esquerda da linha vertical, ao passo que as instituições virtuais, conseguem fornecer toda a gama de artigos: a cauda, bem como os artigos mais procurados. O Fenômeno da Cauda Longa obriga que as instituições *online* recomendem itens para usuários individuais. Não é possível apresentar todos os artigos disponíveis para o usuário, da mesma forma que as instituições físicas conseguem. Nem podemos esperar que os usuários conheçam todos os itens que podem gostar (RAJARAMAN; ULLMAN, 2011).

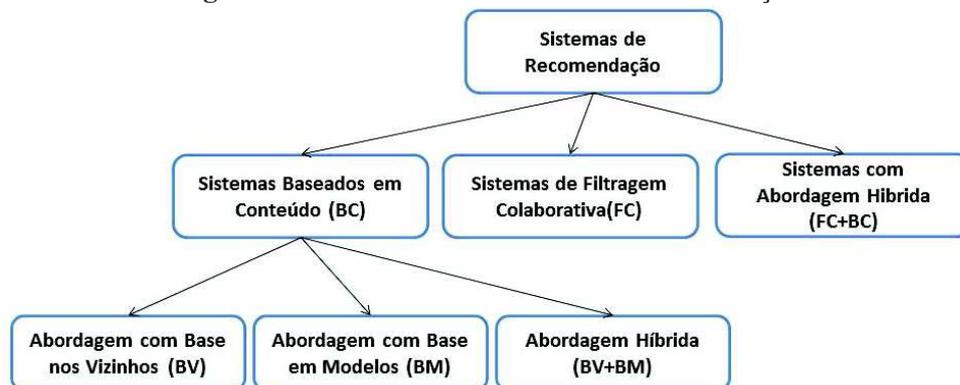
2.6 Sistemas de Recomendação

Existe uma vasta classe de aplicações *web* que envolvem previsão de resposta do usuário para opções. Tal facilidade é denominada de Sistema de Recomendação. Contudo, é necessário se ter algum conhecimento prévio sobre as preferências desse usuário, para que se possa sugerir algo similar ou que tenha relação com esse gosto. Tal cenário resume os dois problemas centrais de qualquer mecanismo de recomendação:

1. Identificação do interesse do usuário, com base na análise do histórico de consumo do mesmo (que itens foram consumidos, quais artigos, textos ou filmes foram avaliados).
2. Cálculo da similaridade entre os itens a serem recomendados, para que se consiga recomendar algo dentro do conjunto de interesses do usuário.

Como solução para este cenário, vislumbra-se hoje duas abordagens distintas de como definir o interesse do indivíduo e como se calcular a similaridade entre os itens, que constituem os dois modelos principais: recomendação por filtragem colaborativa ou baseada em conteúdo. A Figura 3 mostra a taxonomia dos principais sistemas de recomendação existentes.

Figura 3: Taxonomia de Sistemas de Recomendação



Fonte: Adaptada de (RAJARAMAN; ULLMAN, 2011)

- **Sistemas Baseados em Conteúdo (BC):** São sistemas que examinam as propriedades dos itens recomendados. Por exemplo, se um usuário *Netflix*⁴ assistiu muitos filmes de *cowboy*, então lhe é recomendado um filme classificado no banco de dados como tendo o gênero de *cowboy* (RAJARAMAN; ULLMAN, 2011).
- **Sistemas de Filtragem Colaborativa (FC):** Cada usuário e itens possuem propriedades e o sistema recomenda itens com base na similaridade de medidas entre as propriedades dos usuários e/ou itens. Entretanto, a diferença entre esse tipo e o anterior é que aqui o sistema possui a inteligência de recomendar itens desconhecidos para o usuário com base na similaridade das propriedades do usuário/itens. Os itens recomendados para um usuário são os preferidos pelos usuários parecidos (RAJARAMAN; ULLMAN, 2011).
- **Abordagens Híbridas (BC + FC):** Como o nome sugere, são sistemas que mesclam as duas outras abordagens anteriores.

2.7 *Smart Cities*

O conceito de Cidades Inteligente (ou *Smart City*) não está definido exatamente e ainda pode ser considerado como uma ideia vaga. A definição do conceito foi dada por Caragliu em (?) e estendida em Perez-Marínez et.al. em (?) como segue:

"Smart Cities são cidades fundamentalmente criadas com tecnologias da informação e da comunicação (TIC) que investem em capitais humanos e sociais para melhorar a qualidade de vida dos seus cidadãos através do fomento do crescimento econômico, governança participativa, uso dos recursos de forma responsável, sustentabilidade, mobilidade urbana com eficiência, enquanto é garantido a segurança e a privacidade dos seus cidadãos."

Uma *Smart City* pode ser entendida como um ambiente inteligente, que embute TIC que criam ambientes interativos, que trazem a comunicação para o mundo físico. A partir desta perspectiva, uma cidade inteligente (ou em termos mais gerais um espaço inteligente) se refere a um ambiente físico no qual as tecnologias de comunicação e de informação, além de sistemas de sensores, desaparecem à medida que se tornam embutidos nos objetos físicos e nos ambientes nos quais vivemos, viajamos e trabalhamos. A ideia principal é através de sensores e os sistemas de TIC possam trazer mais informações aos cidadãos de forma a tornar a cidade mais produtiva, fazendo um melhor uso de seus recursos, tais como mobilidade urbana, água, energia, serviços públicos e etc.

2.8 *Smart Health*

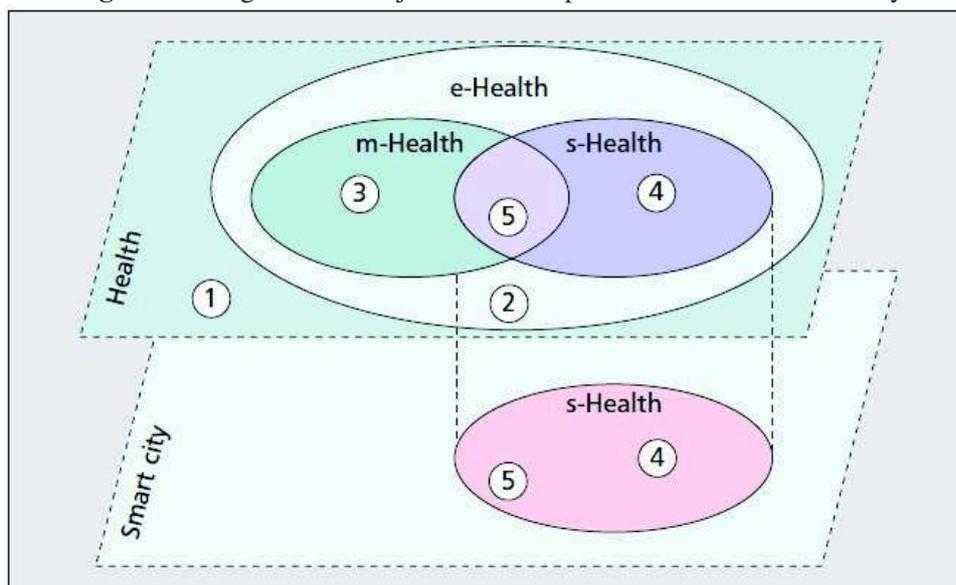
Cunhado por (SOLANAS et al., 2014) o conceito de *Smart Health* (saúde inteligente, em Português) é definido como o complemento natural do *mobile-health* no contexto de *Smart*

⁴www.netflix.com

Cities. A definição, segundo o autor é:

"*Smart Health (S-Health)* é a provisão de serviços de saúde através do uso de uma rede sensível ao contexto usando uma infraestrutura de *Smart Cities*". A Figura 4 ilustra a relação/projeção/intersecção entre os dois planos que é representada como uma área de *S-Health*. A área 1 representa a saúde no geral. A área 2 é o *e-Health*, apresentado na seção 2.4 e a área 3 distingue-se por ser somente via dispositivos móveis.

Figura 4: Diagrama de conjunto entre os planos da saúde e Smart City.



Fonte: Adaptada de (SOLANAS et al., 2014)

2.9 Considerações sobre o Capítulo

A ideia central desta dissertação é apresentar a funcionalidade de recomendação de centros clínicos em uma aplicação *mobile* usando fontes de dados *crowdsourcing* num contexto de *s-health*. Tal funcionalidade seria melhor aproveitada se fosse integrada a um sistema *mobile-PHR* (ou *m-PHR*) num contexto de *Smart City*. Entretanto, a integração do sistema de recomendação com uma solução *m-PHR* é colocada como trabalho futuro.

3 TRABALHOS RELACIONADOS

Os dois primeiros objetivos específico desta dissertação estão relacionados a procurar por trabalhos semelhantes a este e investigá-los. Por isso, este capítulo tem o propósito de apresentar os trabalhos encontrados, como eles foram selecionados, juntamente com uma análise de como os principais trabalhos associados ao tema de pesquisa abordam a questão principal desta proposta. Por fim, os trabalhos são confrontados em uma tabela.

3.1 Metodologia de Pesquisa

A metodologia de pesquisa se deu por meio de buscas utilizando-se de palavras chaves nos mecanismos *onlines* de consulta científica ACM¹, IEEE² e Scholar³ do *Google*. Ela se deu em três etapas distintas entre as quais foi refinando-se a pesquisa.

Na primeira etapa, as palavras chaves buscadas foram *crowdsourcing* e *PHR*. Pois, através das leituras realizadas durante a escolha do tema de pesquisa desta dissertação e também durante as pesquisas feitas para a criação do capítulo de fundamentação teórica acreditou-se que essas seriam as palavras chaves adequadas associadas ao problema em questão. Entretanto, buscando pelas palavras chaves escolhidas nos mecanismos de pesquisa mencionados, foram encontrados poucos resultados: nenhum na IEEE, 4 na ACM e aproximadamente 99 no *Scholar*. Após a leitura de uma parte significativa destes trabalhos, considerou-se as palavras chaves inadequadas. Por isso, refez-se a pesquisa substituindo o termo **PHR** por **Health**, para ampliar os resultados.

A segunda tentativa retornou resultados mais significativos e, por isso, esta foi considerada. Foram levados em consideração aqueles que possuíam a palavra "*crowdsourcing*" e alguma outra palavra que fizesse referência a área de saúde no título do trabalho.

A terceira etapa se deu durante a leitura dos trabalhos encontrados na segunda etapa, onde identificou-se o termo *smart-health* (ou *s-health*). Assim, ele substituiu o termo *health* e uma nova pesquisa foi realizada analisando-se somente esse termo. Em todas as etapas de pesquisa, considerou-se somente publicações recentes e na língua inglesa. Como no período recente, considerou-se a partir do ano de 2012. A seguir são apresentados os trabalhos mais relevantes encontrados durante a segunda e terceira etapa de pesquisa.

3.2 Leveraging The Crowd Collaboration To Monitor The Waiting Time Of Day-To-Day Services

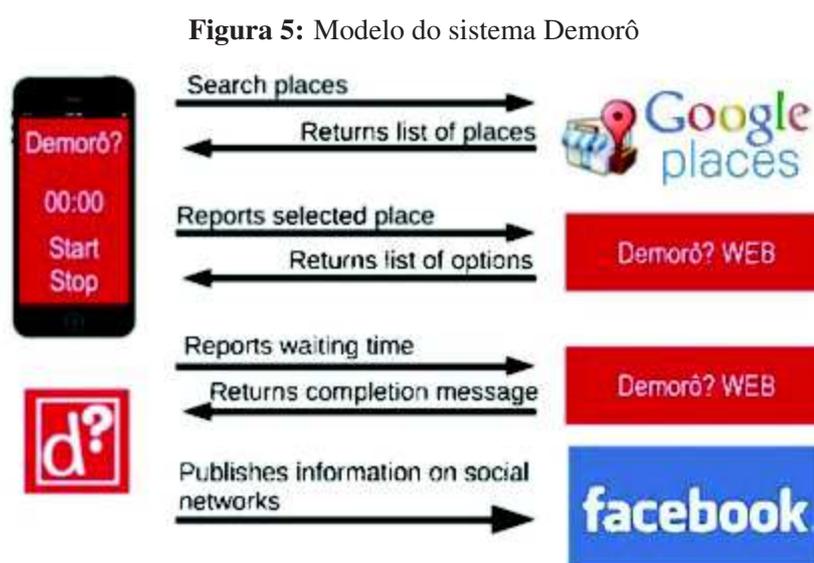
Oliveira apresenta em seu trabalho um aplicativo, chamado Demorô, que tem por objetivo coletar as impressões vividas por usuários para monitorar o tempo de espera e serviço na pres-

¹www.acm.org - Association for Computing Machinery

²ieeexplore.ieee.org - Instituto de Engenheiros Eletricistas e Eletrônicos

³scholar.google.com - Google Acadêmico

tação de serviços públicos e privados. A solução consiste em um ambiente virtual no qual os usuários podem relatar o tempo de espera para a prestação de um determinado serviço. A abordagem usa *crowdsourcing* através dos dispositivos móveis dos usuários e sua capacidade de monitorar a localização dos usuários (OLIVEIRA et al., 2015). A solução é genérica e, segundo os autores, aplicável a diversos tipos de serviços tais como: atendimento ao cliente, serviços bancários, transporte, restaurante, serviços de saúde e etc. A Figura 5 apresenta o modelo do sistema Demorô. O funcionamento do sistema é da seguinte forma:



Fonte: Adaptada de (OLIVEIRA et al., 2015)

1. O usuário faz uma consulta ao sistema que busca na *Google Places API* (API de lugares do Google) por um lugar onde o usuário se encontra e devolve uma lista de opções de lugares.
2. O usuário então seleciona uma opção de serviço e a aplicação Demorô exibe uma lista de serviços oferecidos próximos ao usuário.
3. O usuário informa o tempo de espera para a aplicação. A aplicação informa que os dados foram coletados e questiona se o usuário deseja publicar nas redes sociais.

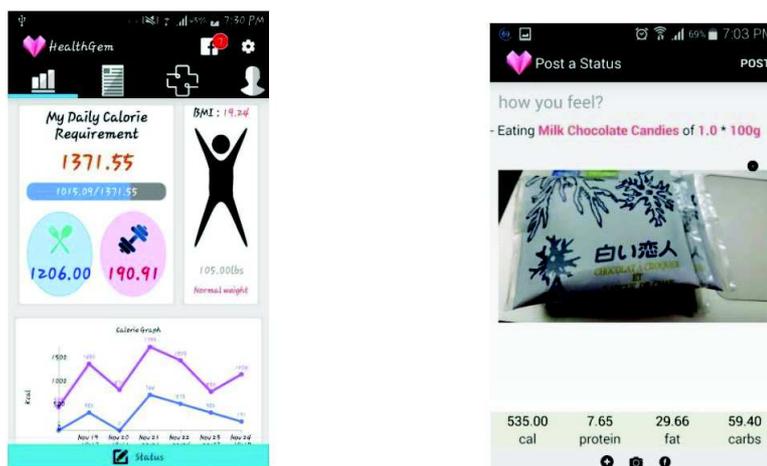
3.3 The Use Of Social Networks In Aid Of Personal Health Monitoring

Anthony apresenta em seu trabalho a proposta de *HealthGem*, que é um PHR com infraestrutura em nuvem que opera com dispositivos móveis *Android* e um servidor *web* em *Java* (ANTHONY; PHILLIP; MAE, 2015). Este servidor *web* atua tanto como *gateway* quanto como *host*. Ele possui diversas características, mas a característica chave é de automaticamente recu-

perar informações de saúde dos usuários através de seus perfis do *Facebook*⁴ usando algoritmos de *data mining*. O público alvo do *app* são usuários que estão interessados em acompanhar sua saúde e seus *posts* regularmente nas redes sociais. A técnica de *data mining* foi a classificação textual através de uma abordagem *N-Gram* (LUO et al., 2013), para análise dos textos das postagens dos usuários.

Em vez de olhar apenas para a cura, o objetivo do aplicativo é a prevenção através do monitoramento e manutenção de um estilo de vida saudável. *HealthGem* possui três módulos principais: *tracker* (rastreador), *display* (exibidor) e *verification* (verificação). O módulo *tracker* é composto por oito rastreadores diferentes que permite aos usuários entrar com informações e rastrear seus registros: alimentação, atividade, peso, pressão arterial, glicemia, exames e notas. O oitavo rastreador é um contador de calorias que calcula ingestão calórica diária do usuário a partir dos rastreadores de alimentos e atividades que informa ao usuário se foram encontradas suas respectivas quantidades calóricas recomendadas. A Figura 6 abaixo exhibe as telas do sistema, a principal e um exemplo de uma postagem feita por um usuário, respectivamente.

Figura 6: Telas do sistema *HealthGem*



(a) Tela Principal

(b) Exemplo de postagem

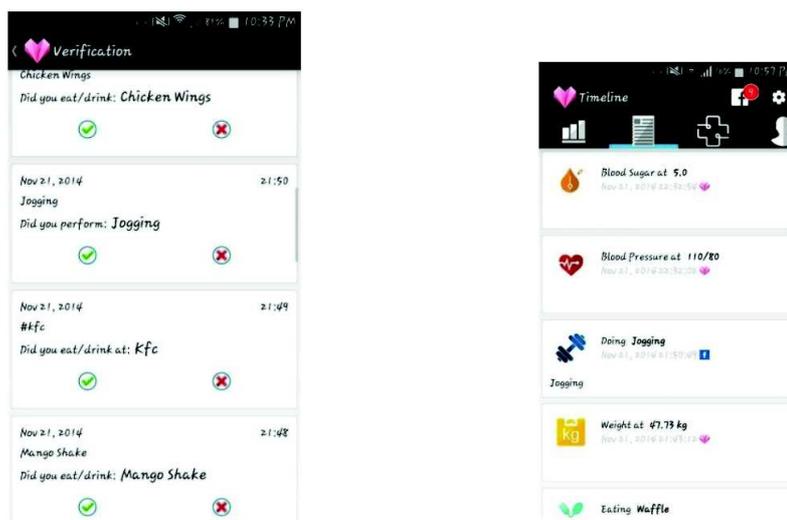
Fonte: Adaptada de (ANTHONY; PHILLIP; MAE, 2015)

O módulo *display* consiste principalmente na linha do tempo, tela inicial, e exibição das informações do módulo *tracker*. O módulo *verification* pergunta ao usuário se ele/ela comeram ou aceitam a informação sobre a saúde recuperada pelo algoritmo automático de filtragem. O algoritmo associa automaticamente o respectivo teor calórico e/ou informação nutricional aos alimentos com as atividades identificadas a partir da conta do usuário do *Facebook*. Para identificação das informações postadas pelos usuários, *HealthGem* conta com o serviço *FatSecret*⁵ para identificação das quantidades calóricas, tanto das atividades quanto dos alimentos, informados pelos usuários. A Figura 7 apresenta as telas do sistema *HealthGem*, a de status do usuário e da linha de tempo, respectivamente.

⁴www.facebook.com

⁵*FatSecret* - www.fatsecret.com ou www.fatsecret.com.br é um *website* que contém uma lista de nomes de atividades com suas correspondências metabólicas.

Figura 7: Telas do sistema *HealthGem*



(a) Exibição de status

(b) Linha do Tempo

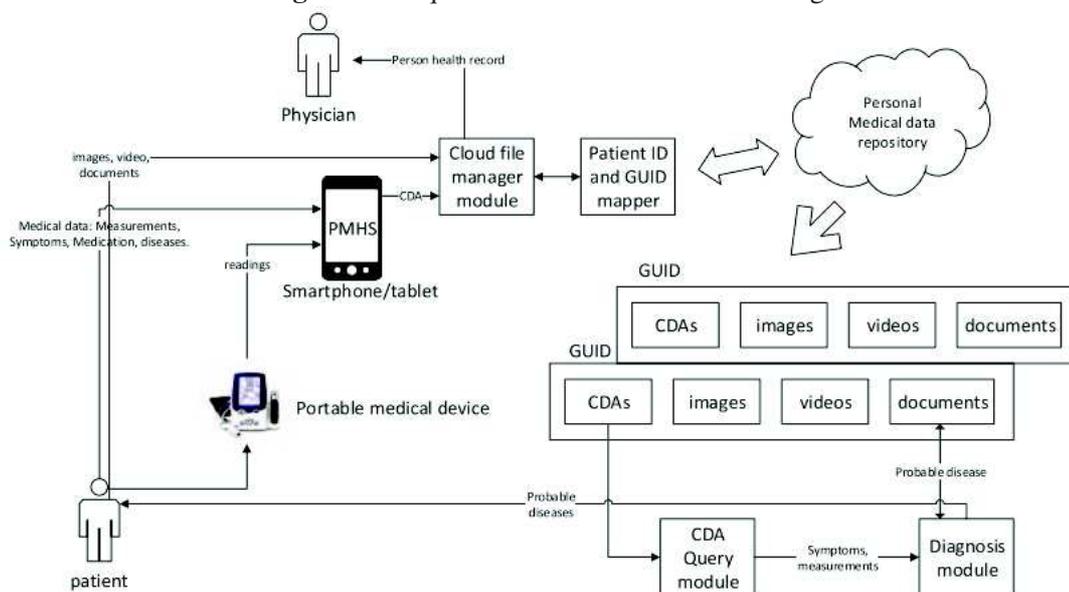
Fonte: Adaptada de (ANTHONY; PHILLIP; MAE, 2015)

3.4 Empowering Patients Using Cloud Based Personal Health Record System

O trabalho de Song apresenta um sistema PHR comum que permite aos usuários monitorar e compartilhar seus dados de saúde com seus médicos. Diferente dos EHR centrados nas instituições clínicas, as entradas dos dados médicos são gerenciadas e controladas individualmente, pelos pacientes ou seus responsáveis. Os PHRs são úteis nos cuidados pessoais, asilos, ou instalações de cuidados privados, onde é necessário controle e monitoramento constante (SONG; HONG; PAK, 2015). Os autores usaram uma aplicação móvel para coletar dados médicos e armazenaram no formato HL7 CDA (EL FADLY et al., 2007) para interoperabilidade. O repositório com infraestrutura em nuvem pode ser compartilhado com os médicos quando necessário. O PHR proposto satisfaz propriedades importantes como acessibilidade, disponibilidade, confiabilidade e confiabilidade. O objetivo do PHR proposto é construir um histórico médico pessoal de longo prazo para os pacientes. A Figura 8 abaixo, apresenta a arquitetura do sistema PHR apresentado no trabalho de Song (SONG; HONG; PAK, 2015). Nela é possível verificar a o fluxo de comunicação das informações dos PHRs entre médicos e usuários.

Em termos de uso significativo, ambos PHR e EHR devem possuir interoperabilidade com cada uma das outras aplicações médicas. Isto é realizado através dos padrões ICD-9-CM (DEYO; CHERKIN; CIOL, 1992), SNOMED CT (DONNELLY, 2006), LOINC (MCDONALD et al., 2003), e HL7 (EL FADLY et al., 2007).

Figura 8: Arquitetura do sistema PHR de Song



Fonte: Adaptado de (SONG; HONG; PAK, 2015)

3.5 CROWDHELP: M-Health App For Emergency Response Improvement Through Crowdsourced And Sensor-Detected Information

No trabalho de Besaleva os autores demonstram um sistema de assistência em tempo real para catástrofes como terremotos, furacões e outros desastre. O sistema usa triagem eletrônica móvel realizada via *crowdsourcing* e informações detectadas por sensores (BESALEVA; WEAVER, 2014).

O trabalho apresenta as aplicações de tecnologias ubíquas, seu o estado da arte nas soluções atuais, questões dos sistemas existentes, a importância das aplicações de redes sociais assim como as tendências futuras. Porém, a arquitetura computacional do sistema não é exibida no trabalho. Segundo os autores, através do uso do sistema, profissionais de gerenciamento de emergência, recebem a maioria das informações que eles necessitam para preparar eles mesmos a realização de um tratamento preciso e oportuno para os pacientes antes mesmo de enviar uma equipe responsável para o local. Durante a demonstração, eles mostram como o sistema se comporta com a combinação de diferentes informações e comparam seus resultados com avaliações feitas por médicos especialistas. O público tem a oportunidade de participar em tempo real provendo informações sobre as vítimas dos acidentes (BESALEVA; WEAVER, 2014).

3.6 PEOPLESERVE: Recommending Effective Drugs Through Web Crowdsourcing

No trabalho de Majethia é apresentado um sistema de análise de revisão médica *crowdsourcing* para a recomendação e interrupção do uso de drogas com base na análise contextual dos comentários das multidões na internet, chamado *PeopleSave*. Os resultados preliminares do estudo, demonstram que existe uma coerência suficiente entre os dados *crowdsourcing* em diferentes domínios geográficos e ontológicos, que é uma métrica desejada para ser capaz de eliminar ou recomendar um subconjunto de drogas a partir de um determinado conjunto de medicamentos disponíveis de forma confiante (MAJETHIA, 2016).

A análise textual foi feita usando uma abordagem *n-gram* inspirada em (CAVNAR; TRENKLE et al., 1994). Após a análise textual, os resultados são submetidos a filtragens realizadas por médicos para verificar se há viabilidade dos dados oriundos de *crowdsourcing* e, então, avaliar a eficácia da droga com base em seu efeito prejudicial relatado em um paciente. Isso ajuda na eliminação de certas drogas que, quase certamente, teriam um efeito adverso sobre a saúde do paciente e, assim, obter um conjunto de drogas recomendáveis. Estas recomendações são ainda mais refinadas através da análise do sentimento detrás das opiniões dos pacientes que foram administrados à estas drogas no passado. O conjunto resultante de drogas prescritas concorda com as sugeridas pelos médicos consultados para o conjunto da amostra considerada de pacientes com diabetes (MAJETHIA, 2016).

A conclusão do estudo é de que a implementação deste sistema pode funcionar como uma ferramenta para os médicos fazerem uso de um grande número de casos anteriores semelhantes como base para tratamento de um doente em particular. Ele irá garantir que cada paciente seja tratado da forma específica no caso mais eficiente possível e, ao mesmo tempo, admitindo o facto de que cada caso é único em relação aos demais (MAJETHIA, 2016).

3.7 SMART FOOD: Crowdsourcing Of Experts In Nutrition And Non-Experts In Identifying Calories Of Meals Using Smartphone As A Potential Tool Contributing To Obesity Prevention And Management

O estudo de Moorhead (MOORHEAD; BOND, 2014) exhibe um sistema para *smartphones* para controle e gerenciamento de dietas alimentares. Os dois principais objetivos do sistema são: determinar a possibilidade de especialistas e não especialistas em nutrição identificar com precisão as quantidades calóricas de refeições através de fotografias obtidas por meio de *smartphones*. O segundo objetivo é desenvolver um sistema de mensagens personalizadas para prevenção e tratamento da obesidade usando um aplicativo móvel.

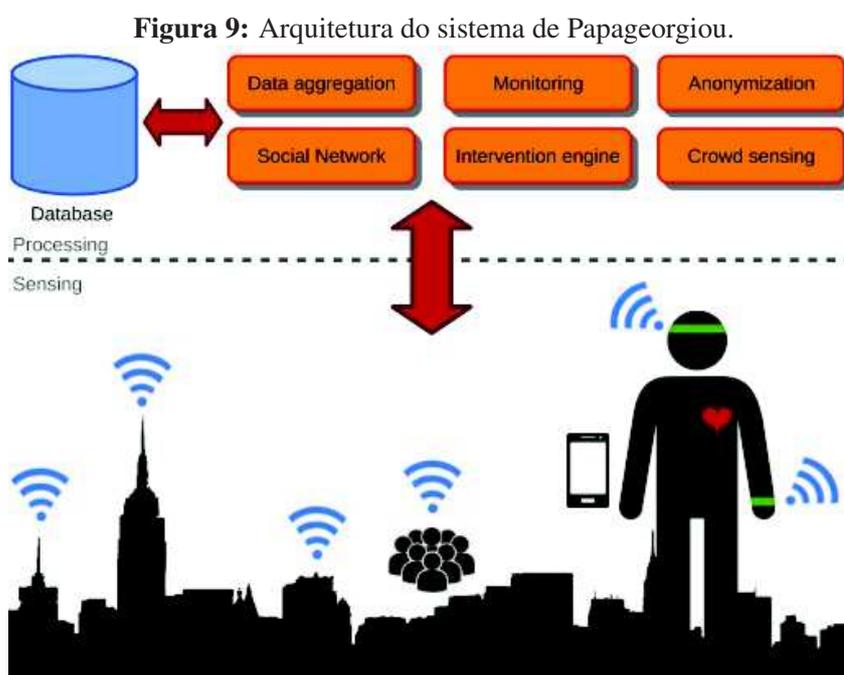
No sistema, os participantes tinham que responder quantas calorias eles achavam que uma dada refeição possuía, apenas olhando para uma fotografia da refeição. O estudo mostrou que os especialistas em nutrição podem estimar com uma precisão de até 8% de erro, enquanto não

especialistas podem fazer a mesma estimativa mas com um percentual de erro de 55%.

3.8 Personalising And Crowdsourcing Stress Management In Urban Environments Via S-Health

No trabalho de Papageorgiou são explorados os potenciais de uma *s-health* para gerenciamento do stress, provendo um *framework* conceitual o qual poderia ser usado não somente para monitorar os níveis de stress, mas também para intervenções mais avançadas de Tecnologia da Informação e Comunicação (PAPAGEORGIU; ZIGOMITROS; PATSAKIS, 2015).

A Figura 9 apresenta o *framework* conceitual assim como seus módulos que são descritos na sequência. A ideia dos autores é de coletar dados de usuários oriundo de *wearables*, sensores urbanos e também fontes *crowdsourcing* para oferecer melhorias de níveis de *stress* no uso de serviços urbanos (PAPAGEORGIU; ZIGOMITROS; PATSAKIS, 2015).



Fonte: Adaptado de (PAPAGEORGIU; ZIGOMITROS; PATSAKIS, 2015)

- **Monitoring**, é responsável por manter o rastreo da localização, movimento e rotina do usuário. A partir dessas medições e das preferências o módulo controla se alguma intervenção é necessária.
- **Data Aggregation** é responsável pela coleta de dados dos sensores da cidade inteligente.
- **Crowd Sensing** é responsável por obter dados de outros usuários. Usuários agindo como sensores podem prover informações adicionais em eventos ou localizações que os sensores comuns não conseguem.

- **Intervention** provê mecanismos pró-ativos quando o sistema detecta que um usuário pode estar numa situação de estresse. Quando um dado nível é atingido o sistema dispara mecanismos que podem ir desde a tocar músicas, fazer novas interações sociais, assistir a notícias desestressantes, tocar vídeos engraçados até exercícios de relaxamento.
- **Social Network** provê informações a respeito das redes sociais, tais como novidades, eventos e interações com outros usuários.
- **Anonymization** é responsável pela desconstrução de qualquer informação que possa levar a identificação exata do usuário.

3.9 FIHUO: A Mobile S-Health Service Platform

O trabalho de Sheng propõe a plataforma *Fihuo* que é um serviço de saúde móvel inteligente com base na análise de *Big Data* (CHEN; ZHANG, 2014). *Fihuo* é capaz de mostrar os dados relacionados a saúde de longo prazo e do ambiente de dados dos usuários. Esses dados são coletados através de dispositivos *wearables*, aplicativos de *smartphones* e mecanismos de *Big Data* rodando em segundo plano (SHENG et al., 2016).

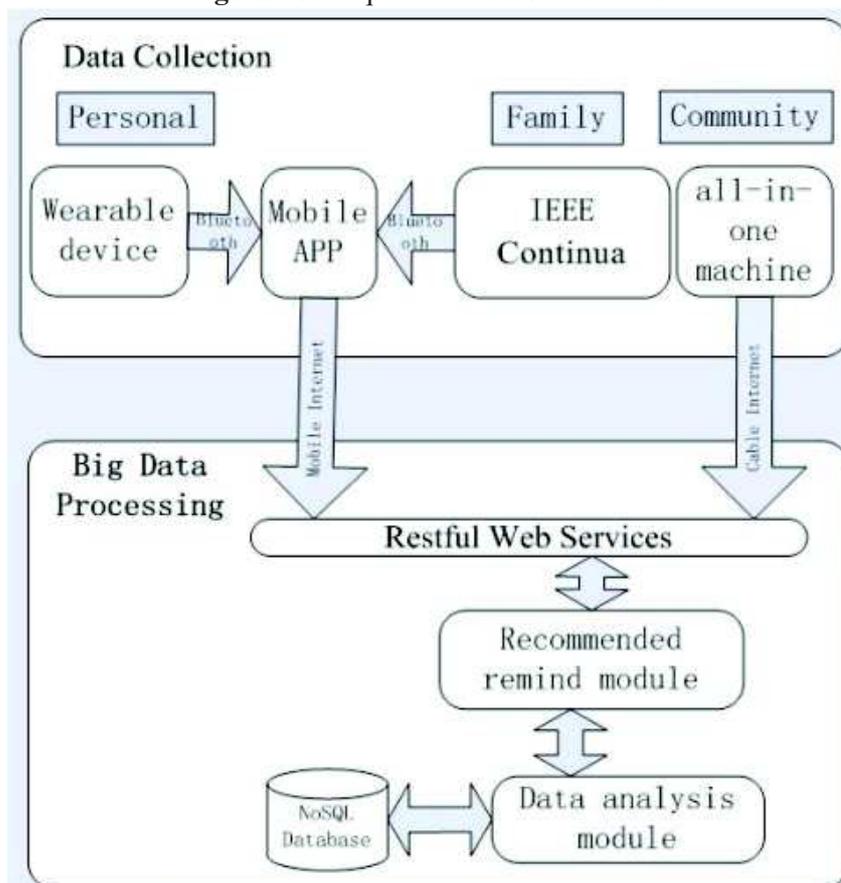
Fihuo foi projetado para poder analisar e comparar dados relacionados a saúde, além de prover um navegador *online* e um monitoramento dos dados de saúde. Os autores desenvolveram também dispositivos *wearables* como pedômetros, *wristband*⁶ e outras aplicações relacionadas a *smartphones*.

Fihuo foi avaliado em usuários que foram classificados em dois grupos. O de doenças crônicas, no caso, diabetes e o de exercícios. O grupo de exercícios tinha como objetivo recomendar atividades físicas para usuários com características semelhantes. A análise da similaridade entre os usuários foi feita usando o cálculo de similaridade do cosseno (CHOWDHURY, 2010). Já o grupo de doenças crônicas tinha o propósito de recomendar refeições de acordo com as características das informações dos dispositivos de saúde dos usuários. A recomendação foi feita usando uma técnica de *deep learning* (aprendizagem profunda, em Português) chamada *self-organized map* (mapa auto-organizável em português). A Figura 10 apresenta a arquitetura do sistema *Fihuo* juntamente com seus módulos de coleta de dados que são descritos a seguir.

- **Wearable Service** provê a coleta de dados pessoais a partir dos dispositivos *wearable* com suporte a tecnologia *Bluetooth*. Isto inclui pedômetros e *wristband* que são usados para coletar dados individuais.
- **Mobile App** provê a coleta de dados dos sensores dos *smartphones* tais como aceleração, peso, luminosidade, câmera, GPS, etc. os quais podem prover funções como treino esportivo, nutrição, pulsação, pressão sanguínea.

⁶espécie de pulseira com dispositivo eletrônico de identificação. É uma tecnologia comum no âmbito de Internet das Coisas

Figura 10: Arquitetura do sistema *Fihuo*.



Fonte: Adaptado de (SHENG et al., 2016)

- **IEEE Continua** Baseado no padrão *IEEE Continua Bluetooth* este módulo inclui monitoramento da pressão sanguínea, oxigênio no sangue, açúcar e batimentos cardíacos,
- **All-in-one machine** é uma plataforma para monitoramento orientado a comunidades multifuncional, e pode monitorar dados do peso, temperatura corporal, pressão sanguínea, pulsação, taxas de oxigênio e açúcar no sangue.

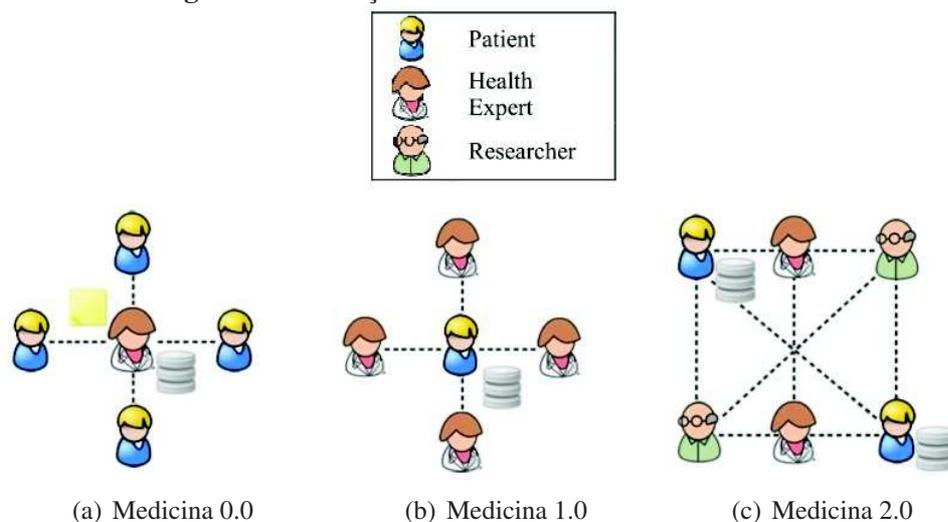
3.10 Comparação entre os Trabalhos Relacionados

Os trabalhos selecionados foram classificados segundo os critérios listados abaixo. Os trabalhos que não possuíam determinada característica informada ou disponível na publicação, ou ainda, a mesma não se aplicava ao trabalho receberam a classificação N/A.

- **Ano:** Ano de publicação do trabalho.
- **Autor:** Nome do autor principal do trabalho.
- **Sistema:** Qual o nome do sistema proposto na publicação.

- **Crowdsourcing:** Se o sistema se utiliza de dados oriundos de fontes *crowdsourcing*, conforme explicado no capítulo anterior.
- **Data Mining:** Se é empregado algum tipo de técnica de *data mining* e qual técnica foi utilizada. Segundo Fayyad, *Data Mining* (Mineração de Dados em Português) pode ser definido por um conjunto de técnicas para analisar dados através de diferentes perspectivas e resumi-los em informações úteis. *Data mining* é atualmente, aplicado em diferentes campos como análise de sentimentos (NASCIMENTO; OSIEK; XEXÉO, 2015), análise de produtos (PEREIRA, 2015) entre outros (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).
- **Tipo de Usuário:** A qual tipo de usuário está destinado a solução proposta.
- **Padrão de Cuidados em Saúde:** Análise do Padrão de Cuidados em Saúde dos trabalhos se deu através da perspectiva de Pereira (PEREIRA et al., 2013). Segundo ele, o modelo de cuidados de saúde vem evoluindo e essa evolução tem alterado o seu foco. O primeiro modelo é o que Pereira chamou de *Medicina 0.0*, que é mais tradicional e centrado na relação médico e paciente. Neste modelo, profissionais de saúde geralmente, usam diferentes abordagens para gerenciar os registros de saúde de seus pacientes, raramente usando recursos de TIC pra isto. Também neste modelo, o paciente somente tem acesso as informações de saúde através de médicos ou agentes de saúde.
A evolução deste modelo é o que Pereira chamou de *Medicina 1.0*, que é centrado no paciente. Neste modelo, são utilizados recursos de TIC, tais como PHR's. O acesso as informações de saúde é feito diretamente pelos pacientes, sem a necessidade de ter a figura do agente de saúde intermediadora. Esse acesso é realizado através de seus PHR's ou ainda usando outros meios como a *Internet* para realizar consultas (PEREIRA et al., 2013).
Por fim, Pereira mostra que a tendencia é o modelo de cuidados de saúde *Medicina 2.0*, onde médicos, enfermeiros e outros profissionais de saúde, pesquisadores de biomedicina, consumidores e pacientes, utilizam aplicações, serviços e ferramentas de infraestrutura *web* que usam *Web 2.0* de forma colaborativa. Os três modelos de cuidados em saúde são ilustrados através da Figura 11
- **Multiplataforma:** Foi avaliado se a solução proposta no trabalho é disponível em mais de uma plataforma para o usuário. Não necessariamente, a mesma aplicação, mas ao menos uma interface de comunicação com o usuário em outro sistema operacional ou outro dispositivo.
- **Objetivo:** Qual é o objetivo da solução proposta no trabalho.
- **PHR:** Se o sistema possui ou utiliza os PHR's.
- **Waiting Time:** Se o sistema monitora, de alguma forma, o tempo de espera dos usuários.

Figura 11: Evolução dos Modelos de cuidados de saúde.



Fonte: Adaptado de (PEREIRA et al., 2013)

Usando os critérios de avaliação apresentados, diversas lacunas de pesquisa de entre os trabalhos podem ser observadas através da Tabela 1. Com relação ao critério foco da solução, é possível observar que nenhuma das soluções propõe algo onde as informações que sejam úteis também à governos ou organizações públicas. E, por fim, mas não menos importante, a questão do modelo de medicina empregado nas soluções, a maioria dos trabalhos ainda não utilizam o modelo de medicina 2.0 sugerido por Pereira que prevê uma relação colaborativa entre médicos, enfermeiros e outros profissionais de saúde, pesquisadores de biomedicina, consumidores e pacientes (PEREIRA et al., 2013). Os trabalhos apresentados na área de saúde não exploram a questão de utilizar dados oriundos de fontes *crowdsourcing* e, em troca, propor diminuição no tempo de espera de serviços de saúde para os usuários. Também não foi encontrado nenhum trabalho que faça uma recomendação acerca da escolha mais útil.

É possível observar também que os PHR's estão presentes na maioria das publicações selecionados. Assim, é identificada uma lacuna de pesquisa que toque a questão do foco também a entidades governamentais ou organizações públicas, que possua um modelo de medicina onde prevendo uma relação colaborativa entre os profissionais da saúde, consumidores e pacientes. Que toque também a questão do uso de *crowdsourcing* como colaboração no auxílio da tomada de decisão acerca de qual centro clínico é mais útil para o usuário. Essa lacuna de pesquisa será explorada no modelo de solução do *CrowdHealth* apresentado no capítulo a seguir.

Tabela 1: Comparação entre trabalhos correlatos

Sistema	Demorô	HealthGem	N/A	PeopleSave	Smart Food	N/A	FIHUO	CrowdHelp
Ano	2015	2015	2015	2016	2014	2015	2016	2014
Autor Principal	Oliveira	Anthony	Song	Majethia	Moorhead	Papageorgiou	Sheng	Besaleva
Crowdsourcing	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim
Data Mining	N/A	Classificação Textual (N-Gram)	N/A	Classificação Textual (N-Gram)	N/A	N/A	Similaridade do Cosseno & self-organized map	N/A
Foco	Cidadãos	Paciente	Pacientes e médicos	Pacientes e médicos	Pacientes e Agentes de saúde	Cidadãos	Cidadãos	Cidadãos
Multiplataforma	Sim	Não	Não	Não	Sim	Sim	Sim	Sim
Objetivo	Mapear tempo de espera de serviços de uma cidade	Melhoria de hábitos de saúde através de compartilhamento de informações.	Construir um histórico médico pessoal de longo prazo para os pacientes.	Análise de revisão médica <i>crowdsourcing</i> para a recomendação e interrupção do uso de drogas.	Controle e gerenciamento de dietas alimentares.	Gerenciamento do estresse em cidades inteligentes.	Análise e comparação de dados relacionados a saúde	Assistência em tempo real para catástrofes como terremotos, furacões e outros desastre
PHR	N/A	Sim	Sim	Sim	Sim	Sim	Sim	Não
Padrão de Cuidados em Saúde	N/A	1.0	0.0	0.0	2.0	1.0	1.0	1.0
Wait Time	Sim	Não	Não	Não	Não	Não	Não	Não

Fonte: Elaborado pelo autor.

4 CROWDHEALTH

O terceiro objetivo específico desta dissertação é apresentar um modelo científico de solução que motive multidões a compartilhar seus dados de saúde pessoais por meio de uma relação ganha-ganha entre os usuários envolvidos. Assim, este capítulo apresenta o modelo do *CrowdHealth*, que consiste basicamente na criação desta relação. Na seção seguinte, são apresentados a visão geral do modelo e seus objetivos. Já as táticas usadas durante sua criação são explicadas na seção 4.2. A seção 4.3 apresenta a formalização do problema de recomendação de centros clínicos de saúde para usuários. Por fim, a arquitetura do modelo é apresentada na seção 4.4.

4.1 Premissas

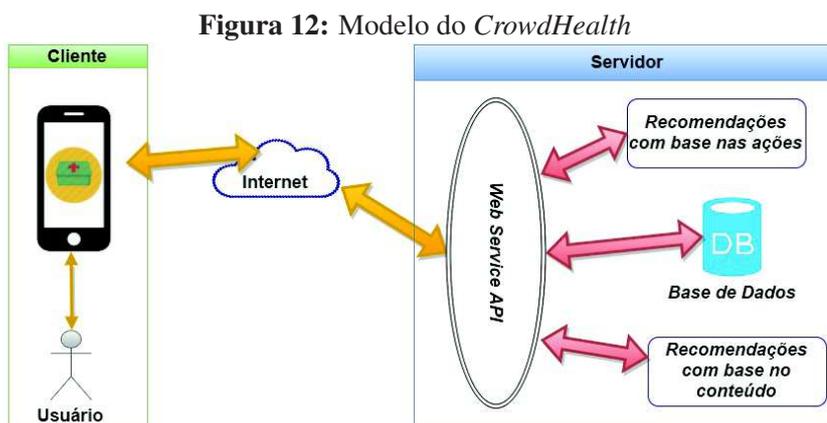
O objetivo do *CrowdHealth* é propor uma relação ganha-ganha entre pacientes, profissionais de saúde e órgãos públicos. Essa relação acontece no compartilhamento de informações privadas do usuário e relacionadas a qualidade do seu atendimento. Com relação as informações do usuário, este compartilha sua localização e seus dados de saúde. E em relação a seu atendimento, o usuário informa o local onde recebeu o atendimento, o tempo de atendimento e também qualifica o atendimento recebido. Ao compartilhar os seus registros, o usuário tem acesso a informações do registro de todos. Isso é análogo a aplicação de dispositivos móveis Waze¹, onde usuários fornecem informações de geolocalização em tempo real, mas que não seria do interesse deles compartilhar sem receber algo em troca, no caso o status do fluxo de veículos nas ruas da cidade.

O modelo do *CrowdHealth* foi inspirado no trabalho de Oliveira, porém especificamente na área da saúde (OLIVEIRA et al., 2015). A Figura 12 ilustra o modelo do *CrowdHealth*. De forma mais macro, *crowdhealth* possui apenas uma aplicação que roda nos *smartphones* dos usuários, uma base de dados e módulos de recomendação com base no conteúdo e também nas ações dos usuários. O módulo com base no conteúdo serve para evitar o problema de partida a frio. E o módulo de recomendações com base nas ações serve pra criar as recomendações do tipo filtro colaborativo apresentado na seção 2.6.

Toda comunicação entre a aplicação que roda nos *smartphones* dos usuários e o sistema acontece através do *Web Service*. *Web Service* é uma solução utilizada na integração de sistemas e na comunicação entre aplicações diferentes. Com esta tecnologia é possível que novas aplicações possam interagir com aquelas que já existem e que sistemas desenvolvidos em plataformas diferentes sejam compatíveis. Os *Web Services* são componentes que permitem às aplicações enviar e receber dados em formato XML. Cada aplicação pode ter a sua própria "linguagem", que é traduzida para uma linguagem universal, um formato intermediário como XML, JSON, CSV, etc. Os usuários interagem com seus *smartphones* e este com o *web service*. Ao necessitar

¹<https://www.waze.com>

de um centro de saúde, o usuário acessa o *CrowdHealth* em busca de qual centro de saúde é o mais aconselhável que ele vá. Assim, através do *smartphone* o usuário informa a localização atual e o serviço de saúde desejado. O *CrowdHealth* retorna uma lista de opções de centros clínicos ordenadas com base nas características do usuário e suas preferências. O usuário visualiza as opções e seleciona um centro clínico desejado. Nas opções são exibidos os nomes, localização, tempo médio de atendimento e avaliação média dos centros clínicos.



Fonte: Elaborado pelo Autor

Através do sensor de geolocalização do *smartphone* onde está aplicação do usuário, o sistema monitora a chegada do usuário ao local. O *CrowdHealth* obtém as preferências do usuário através do histórico de suas ações. Quais os centros clínicos que o usuário já visualizou, desses quais utilizou, e ainda, quais os que avaliou positivamente e o contexto em que essas ações aconteceram. O contexto aqui é o momento e a localização em que o usuário estava.

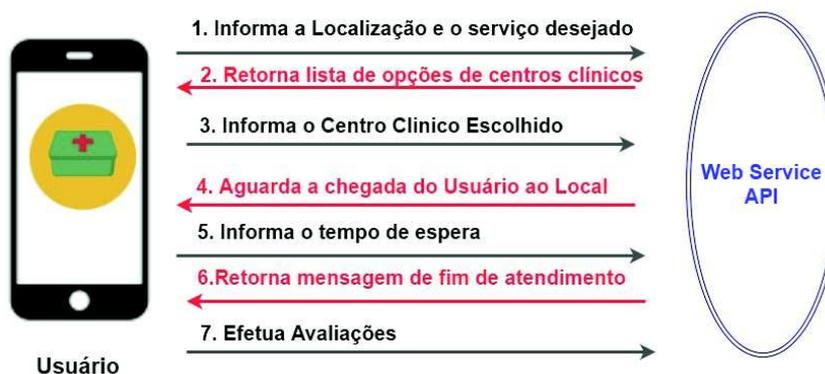
Ao deixar local ou finalizar o atendimento, o usuário informa o tempo de atendimento no centro clínico. Ao fazê-lo o *CrowdHealth* retorna mensagem de finalização de atendimento. Neste momento então, o usuário efetua avaliações com relação ao atendimento recebido no centro clínico e a recomendação de centros clínicos oferecidas. Também neste momento é que o usuário autoriza o compartilhamento de seus dados de saúde.

A Figura 13 ilustra a interação do *CrowdHealth* com o usuário.

1. O usuário informa a localização atual e o serviço de saúde desejado.
2. O sistema retorna uma lista de opções de centros clínicos ordenadas com base nas características do usuário e suas preferências.
3. O usuário informa o centro clínico escolhido.
4. O sistema, através do sensor de geolocalização do *smartphone* onde está aplicação do usuário, monitora a chegada do usuário ao local.
5. O usuário informa o tempo de atendimento no centro clínico previamente escolhido.

6. O sistema retorna mensagem de finalização de atendimento.
7. O usuário efetua avaliações com relação ao atendimento recebido no centro clínico, a recomendação de centros clínicos oferecidas e também autoriza o compartilhamento de seus dados de saúde.

Figura 13: Comunicação do *CrowdHealth* com os usuários



Fonte: Elaborado pelo Autor

4.2 Estratégias e Decisões de Projeto

A parte mais importante de um sistema de recomendação é o mecanismo que provê as recomendações. Sistemas de recomendação podem ser do tipo baseados em conteúdo, filtros colaborativos ou abordagens híbridas, conforme apresentado na seção 2.6 do Capítulo 2. Recomendações com base em conteúdo possuem limitações no sentido de não recomendar itens novos, ou itens ainda desconhecidos para o usuário. Por outro lado, recomendações usando mecanismos de filtragens colaborativas, também conhecido como "quem gostou disso, também gostará daquilo", apesar de suprirem a carência do tipo anterior, sofrem de um problema conhecido na literatura como "partida a frio"². Daí surge a necessidade de utilização de sistemas de recomendação com abordagens híbridas. Por isso, a arquitetura e o modelo do *CrowdHealth* foi planejada pensando nessa necessidade, conforme poderá ser visto na seção 4.4.

O cálculo de tempo médio de atendimento e média de avaliações dos centros clínicos independe do usuário e também são rotinas demoradas, visto que calcular todos os registros informados para cada local é uma tarefa complexa. Por este motivo, optou-se por dividir as rotinas de cálculo de acordo com o tempo de processamento. Isto é importante para que, para o usuário, pareça que a recomendação é calculada em tempo real, quando na verdade, o cálculo de tempo

²dificuldade do sistema em fazer recomendações para o usuário na etapa inicial, quando o sistema ainda não possui informações suficientes do usuário e suas preferências

médio de atendimento e a média das avaliações dos centros clínicos são processadas de maneira assíncrona. Outra decisão importante foi a questão de desvalorizar significativamente as distâncias superiores a distância informada pelo usuário, bem como as avaliações de acordo com a atualidade da informação. A importância dessas decisões são observadas no capítulo seguinte, na seção de resultados.

4.3 Formalização do Problema

Seja U o conjunto de todos os usuários do sistema, e seja C' o conjunto de todos os possíveis centros clínicos de saúde que podem ser recomendados. Seja a a função adequação que mede o quão adequado é um determinado item c para um determinado usuário u . Por exemplo, $a : U \times C \rightarrow R$, onde R é um conjunto totalmente ordenado. Então, para cada usuário $u \in U$, procura-se um item $c' \in C$ que maximiza a utilidade do usuário (DEBNATH; GANGULY; MITRA, 2008). A Equação 4.1 abaixo expressa formalmente essa questão:

$$\forall u \in U, c'_u = \operatorname{argmax}_{c \in C} a(u, c) \quad (4.1)$$

Assim, a função adequação (a) de um usuário (u) em relação a um centro clínico (c) leva em consideração três fatores: a Distância ($Dist$), o Tempo Médio de Atendimento ($ServiceTime$) e as avaliações ($Rating$). Abaixo segue a definição de cada termo.

- $Dist_{(u,c)}$ é a distância de um usuário u a até um centro de clínico c .
- ΔT é a diferença entre as marcas temporais do momento que o usuário faz a consulta e o momento que uma informação foi gravada.
- $Dist_{(d)}$ é a distância que o usuário está disposto a percorrer para encontrar o centro clínico em Kilometros.
- $ServiceTime_{(u,c)}$ é o tempo de atendimento informado por um usuário para um centro clínico.
- $\overline{ServiceTime}_{(c)}$ é a média aritmética ponderada dos tempos de espera informados pelos outros usuários para um centro clínico com relação ao momento informado. Quanto mais recente a informação, maior será seu peso na composição da média do tempo de espera.
- $Rating_{(u,c)}$ é a avaliação de um usuário para um determinado centro clínico.
- $\overline{Rating}_{(c)}$ é a média aritmética ponderada das avaliações que os usuários deram ao atendimento que receberam em um centro clínico.
- $\%DesuWaitingTime$ é o percentual de desvalorização do tempo de espera ao longo do tempo.

- $\%DesvRating$ é o percentual de desvalorização das avaliações ao longo do tempo.

Considerando-se a capacidade de locomoção média de uma pessoa normal pode variar, optou-se por utilizar uma heurística que desvalorizasse as distâncias numa escala de $Dist_{(d)}$ Km, ou seja, a cada $Dist_{(d)}$ Km a recomendação de um determinado centro clínico recebe uma desvalorização maior. Isto serve para que a recomendação de um centro clínico mais próximo, receba maior destaque, mesmo nas situações em que a sua avaliação não é tão boa, ou quando o tempo de espera não é baixo. Dessa forma, o cálculo para recomendação de centros clínicos para um usuário através da Distância ($Dist$) desse mesmo usuário até o determinado centro clínico é expresso através da Equação 4.2.

$$Dist_{(u,c)} = \left| \frac{d_{(u,c)}}{Dist_{(d)}} \times d_{(u,c)} \right| \times -1 \quad (4.2)$$

Inspirado no Fenômeno da Cauda Longa, apresentado na seção 2.5 do capítulo Fundamentação Teórica, o tempo de atendimento e as avaliações são calculados através de uma média aritmética ponderada, que leva em consideração o quão recente é cada informação. A medida que o tempo passa, menos importante se torna um dado de tempo de atendimento informado por um usuário para aquele local. Desta forma, o cálculo do Tempo Médio de Atendimento de um determinado centro clínico é expresso através da Equação 4.3, ao passo que as avaliações são expressas na Equação 4.4.

$$\overline{ServiceTime}_c = \sum_1^{Qtd.Avaliacoes} \times \frac{ServiceTime_{(u,c)} \times \Delta T \times \%DesvDesvWaitingTime}{Qtd.Avaliacoes} \quad (4.3)$$

$$\overline{Rating}_c = \sum_1^{Qtd.Checkin's} \times \frac{Nota \times \Delta T \times \%DesvRating}{Qtd.Checkin's} \quad (4.4)$$

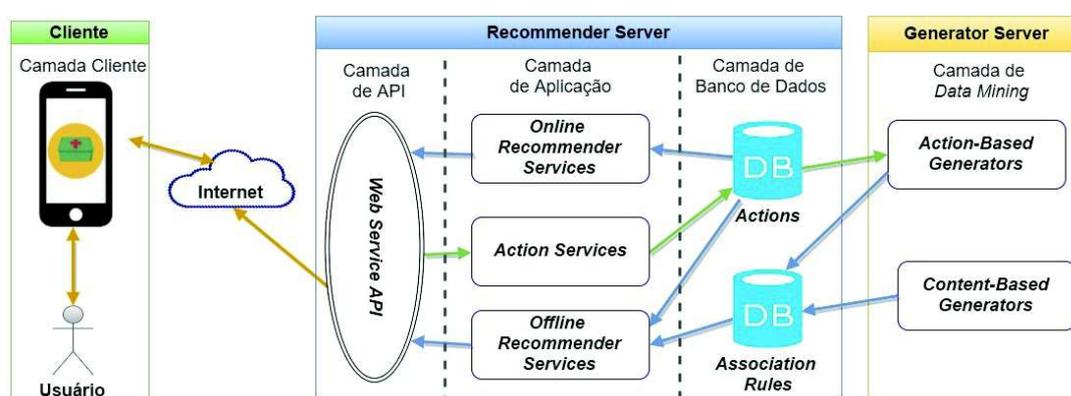
4.4 Arquitetura

Segundo a Microsoft, uma arquitetura, deve ser baseada em componentes, centrando-se na decomposição do projeto em componentes funcionais ou lógicos, divididos em camadas, possuindo como principais características a alta coesão e o fraco acoplamento, favorecendo a evolução da aplicação com novas camadas (Microsoft Corporation, 2009). A arquitetura proposta do *CrowdHealth* deve refletir as características mencionadas pela Microsoft, possuindo cinco camadas. As camadas de *API*, *Aplicação*, *Banco de Dados* e *Data Mining* ficam do lado do servidor e enquanto que a camada *Cliente*, fica do lado do cliente. A Figura 14 demonstra a arquitetura do *CrowdHealth* e suas camadas são descritas a seguir:

- A Camada do Cliente é responsável pela aplicação que roda nos *smartphones* dos usuários.

- A Camada API (*Application Programming Interface*) prove a interface de comunicação da aplicação dos clientes com as rotinas da camada de Aplicação.
- A Camada de Aplicação é responsável por receber as requisições dos usuários e responder através de recomendações.
- A Camada de Banco de Dados é responsável pelo armazenamento das informações das ações do usuário e o resultado das rotinas pré-processadas das regras de negócio
- A Camada de *Data Mining* é responsável pelas rotinas de geração de conhecimento.

Figura 14: Arquitetura do *CrowdHealth*



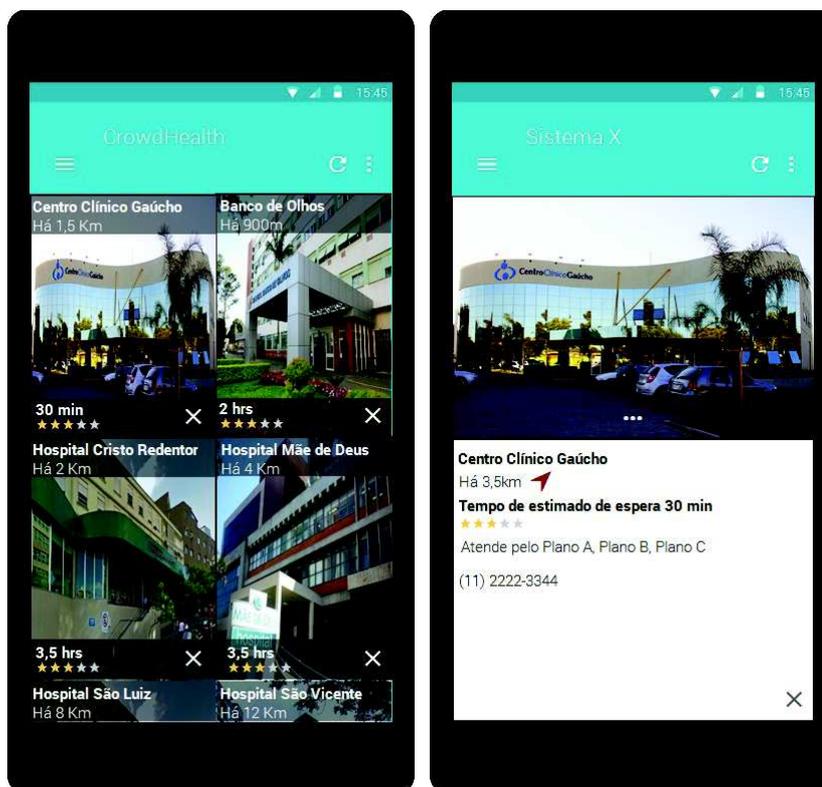
Fonte: Elaborado pelo Autor

4.4.1 Camada do Cliente

É responsável pela aplicação que executa nos *smartphones* dos usuários. A Figura 15 apresenta o protótipo da aplicação. A tela de Visualização é a funcionalidade principal. Nela são exibidos todas os centros clínicos recomendados para o usuário. A tela de Detalhe, aparece quando o usuário seleciona uma recomendação. Nela são exibidas algumas informações adicionais a respeito do centro clínico selecionado.

A Figura 16 exibe outras duas telas importantes do *CrowdHealth*: A tela de Cronometro (a) e a tela de Mensagem. A tela de Cronometro aparece quando *CrowdHealth* identifica através do sensor de geolocalização do *smartphone* onde está aplicação do usuário e se este já chegou ao centro clínico de saúde. Essa tela oferece ao usuário a opção dele começar a cronometrar o seu tempo de espera. Por fim, a tela de Mensagem exibe a mensagem a finalização do tempo de atendimento recebido pelo usuário. *CrowdHealth* coleta esse tempo através dos dados de geolocalização do sensor do *smartphone* do usuário, do momento em que ele iniciou a contagem do tempo de espera até o momento em que saiu. Ela oferece a opção do usuário qualificar o atendimento que recebeu no centro clínico e também a opção de compartilhar o seu tempo de atendimento nas redes sociais para incentivar o uso do aplicativo por novos usuários.

Figura 15: Protótipo da Aplicação do Cliente



(a) Visualização

(b) Detalhe

Fonte: Elaborado pelo autor.

4.4.2 Camada de API

A camada de API (Interface de Processamento da Aplicação, do inglês *Application Process Interface*) possui como principal componente o *Web Service*. Ela é responsável pela interface de comunicação entre as aplicações que rodam nos *smarthphones* dos clientes e as rotinas da camada de Aplicação. As interações entre aplicações *mobile* e *web-service* podem usar dois tipos de protocolos de comunicação, JSON³ ou XML⁴.

A Tabela 2 apresenta uma descrição dos serviços disponibilizados pelo *Web Service*. Os serviços são como funções que possuem parâmetros de entrada, executam algum processamento e devolvem uma saída.

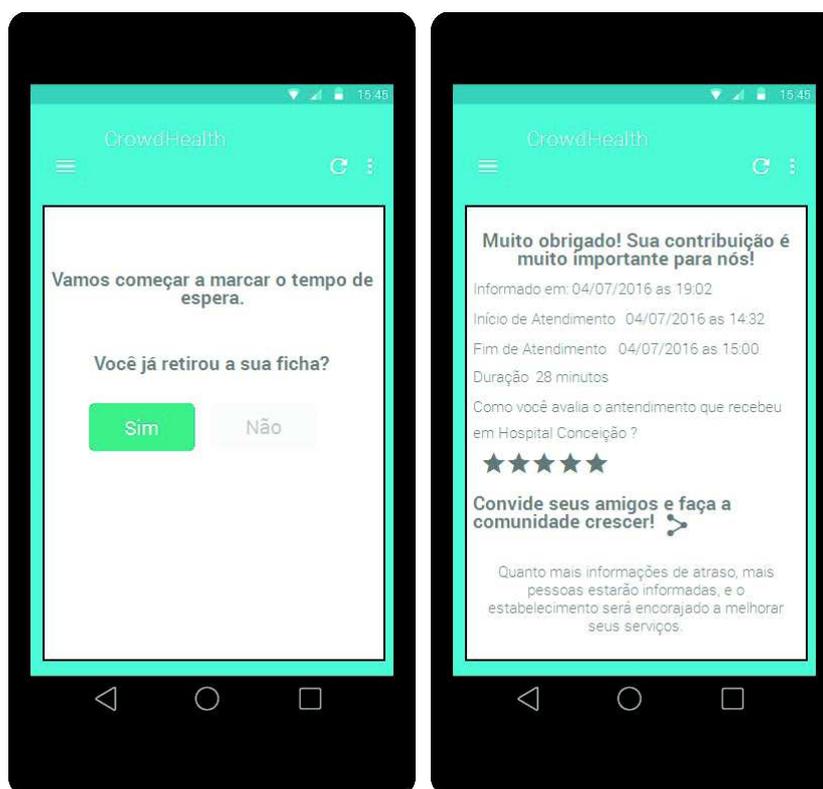
4.4.3 Camada de Aplicação

A Camada de Aplicação é responsável por receber as requisições dos usuários e responder através de recomendações. O processamento das recomendações foi dividido em dois componentes de acordo com o momento em que são processados. As rotinas que são processadas em

³acrônimo de Notação de Objetos *JavaScript* do inglês *JavaScript Object Notation*

⁴acrônimo de *eXtensible Markup Language* ou, em português Linguagem Extensível de Marcação

Figura 16: Protótipo da Aplicação do Cliente



(a) Cronometro

(b) Mensagem

Fonte: Elaborado pelo autor.

tempo real ficam no Serviço de Recomendação *Online*, ao passo que as rotinas que são processadas de forma assíncrona ficam no Serviço de Recomendação *Offline*. O *Action Service* é o serviço responsável por persistir as ações do usuário na tabela *Actions*.

4.4.4 Camada de Banco de Dados

A camada de Banco de Dados possui como principais funções o armazenamento e persistência das informações referente as ações do usuário e o resultado das rotinas pré-processadas das regras de negócio. Ela possui duas tabelas principais *Actions* e *Association Rules*. A tabela *Action* guarda as informações relacionadas as das ações do usuário. A tabela *Association Rules* guarda as regras de associação.

4.4.5 Camada de *Data Mining*

A Camada de *Data Mining* provê as rotinas de geração e extração de conhecimento. Ela possui dois módulos o *Action Based Generators* e o *Content Based Generators*. O módulo *Action Based Generators* é responsável pela mineração de dados para geração de conhecimento com base nas ações dos usuário. O módulo *Content Based Generators* é responsável pela

Tabela 2: Serviços oferecidos pelo componente *Web Service*

Nome do Serviço	Parâmetros de Entrada	Processamento	Parâmetros de Saída
<i>login</i>	O usuário, a senha e a versão da aplicação do cliente.	Realiza o início de uma sessão.	Devolve, em caso de sucesso, o identificador do utilizador e uma chave da sessão iniciada.
<i>logout</i>	Identificador do utilizador e a chave da sessão.	Finaliza/destrói a sessão.	Devolve, em caso de sucesso 1, senão 0.
<i>setLocation</i>	Identificador do utilizador, a chave da sessão e as coordenadas de localização do usuário.	Grava as informações de latitude e a longitude.	Devolve, em caso de sucesso 1, senão 0.
<i>getHealthService</i>	Identificador do utilizador e a chave da sessão.	Busca no sistema uma lista de serviços de saúde disponíveis.	Uma lista de identificadores e nomes de serviços de saúde.
<i>setHealthService</i>	Identificador do utilizador, a chave da sessão e o identificador de um serviço de saúde.	Grava o identificador de serviço de saúde para aquele usuário naquela sessão.	Devolve, em caso de sucesso 1, senão 0.
<i>getClinicalCenter</i>	Identificador do utilizador, a chave da sessão.	Busca no sistema uma lista de centros clínicos próximos ao usuário e que disponham do serviço de saúde desejado.	Devolve uma lista de identificadores e nomes de centros de saúde, ordenadas conforme a recomendação.
<i>setClinicalCenter</i>	Identificador do utilizador, a chave da sessão e o identificador do centro de saúde.	Grava no sistema o identificador do centro de saúde escolhido pelo usuário.	Devolve, em caso de sucesso 1, senão 0.
<i>setPrintClinicalCenter</i>	Identificador do utilizador, a chave da sessão e o identificador do centro de saúde.	Grava no sistema o identificador do centro de saúde visualizado pelo usuário.	Devolve, em caso de sucesso 1, senão 0.
<i>setWaitStart</i>	Identificador do utilizador, a chave da sessão.	Grava no sistema a marca temporal em que o usuário começou a esperar.	Devolve, em caso de sucesso 1, senão 0.
<i>setEndWait</i>	Identificador do utilizador, a chave da sessão.	Grava no sistema a marca temporal em que o usuário finalizou a espera.	Devolve, em caso de sucesso mensagem de agradecimento, senão 0.
<i>setRecommendation Review</i>	Identificador do utilizador, a chave da sessão e uma nota (valor de 1 a 5).	Grava no sistema a nota que o usuário deu para a recomendação recebida naquela sessão.	Devolve, em caso de sucesso mensagem de agradecimento, senão 0.
<i>setClinicalCenterReview</i>	Identificador do utilizador, a chave da sessão e uma nota (valor de 1 a 5).	Grava no sistema a nota que o usuário deu para o atendimento recebido naquele centro clínico.	Devolve, em caso de sucesso mensagem de agradecimento, senão 0.

Fonte: Elaborado pelo Autor

mineração de dados para geração de conhecimento com base nos perfis dos usuários. Como *CrowdHealth* funciona de maneira colaborativa no perfil do usuário é avaliado a sua reputação para que somente usuários com boa reputação, ou se de prioridade para as informações obtidas através de usuários com boa reputação.

5 METODOLOGIA DE AVALIAÇÃO E RESULTADOS

Neste capítulo é mostrado como o *CrowdHealth* foi avaliado. A avaliação é importante para garantir a efetividade dos algoritmos de sistemas de recomendação. Assim, na seção 5.1 é mostrado algumas decisões importantes que foram feitas para avaliação. Já na seção 5.2, é descrito os é mostrado os objetivos e critérios de avaliação. Na sequência, na seção 5.3, é demonstrado como se deu a implementação do modelo do *CrowdHealth*. Logo depois, na seção 5.5, é descrito como era a arquitetura do ambiente de testes. Em seguida, na seção 5.4 é demonstrado como foi calculado o método de avaliação *as Cegas*. Por fim, na seção 5.6, são mostrados os resultados da avaliação.

5.1 Decisões de Projeto

Os percentuais de desvalorização das informações de Tempo de Espera e Avaliação são informações que o sistema só poderia encontrar se a base de dados já estivesse populada. O valor da distância que o usuário está disposto a percorrer, também é uma informação que diz respeito a cada usuário em particular. Para dar seguimento nas avaliações algumas decisões neste sentido foram tomadas. Nas fórmulas de recomendação por Tempo de Espera e Avaliação, optou-se por desvalorizar a informação do tempo de atendimento numa medida três vezes superior em relação as avaliações dos usuários. Desta forma foi utilizado o valor de 0,01 para a desvalorização da informação do atendimento e 0,03 para a desvalorização da informação do Tempo de espera. Optou-se por escolher como cinco Km para o valor de $Dist_{(d)}$ que representa a distância que o usuário está disposto a percorrer para encontrar o centro clínico.

5.2 Critérios de Avaliação

O objetivo principal da avaliação de um sistema de recomendação é saber se o mecanismo de recomendação funciona e, em seguida, saber o quão bom é o seu desempenho. Portanto, o objetivo na avaliação do *CrowdHealth* é saber se os itens (os centros clínicos) oferecidos aos usuários são os mais úteis dentro do contexto de cada usuário. O contexto de cada usuário inclui o momento em que ele faz a requisição de recomendação ao sistema e a localização onde ele se encontra.

Existem diversas métricas estatísticas para avaliação de algoritmos de sistemas de recomendação, tais como Erro Quadrático Médio (EQM, ou MSE em inglês), Desvio Médio da Raiz Quadrada (DMRQ ou RMSD em inglês), *Click-Through Rate* (CTR) (RAJARAMAN; ULLMAN, 2011), *precision* (precisão) e *recall* (recuperação). As métricas *precision* e *recall* são métricas de aferição de recuperação da informação, elas são úteis para garantir a qualidade do método de recomendação (GUNAWARDANA; SHANI, 2009).

Segundo Gunawardana, para a tarefa de recomendar itens, geralmente o interesse recai sob

taxas binárias, isto é, ou o item foi selecionado (1) ou não (0). Comparar taxas em *datasets*, onde usuários tipicamente avaliam somente um pequeno número de itens cria conjuntos de dados extremamente esparsos. Selecionar binários é uma tarefa densa, visto que cada item foi ou não selecionado pelo usuário. Um exemplo de como os conjuntos de dados são fluxos de históricos de visualizações onde, montamos um valor 1 para cada item que foi visitado, e um valor 0 para cada um dos que não foram. A tarefa é prover, dado uma lista de itens existentes, quais foram visualizados, uma lista de itens adicionais que o usuário pode desejar ver (GUNAWARDANA; SHANI, 2009). Como foi explicado acima, esse cenário é tipicamente assimétrico. Nós não estamos igualmente interessados em itens bons ou ruins. Nós podemos classificar os resultados como recomendações usando a Tabela 3.

Tabela 3: Métricas de Avaliação

	Recomendados	Não Recomendados
Preferidos	Verdadeiros Positivos (VP)	Falso-Negativo (FN)
Não preferidos	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Adaptado de (GUNAWARDANA; SHANI, 2009).

Portanto, podemos agora contar qual o número de exemplos que se encaixam em cada célula da tabela e calcular isso usando as métricas *Precision* e *Recall* (Taxa de Verdadeiro Positivos), conforme apresentado pelas Equações 5.1 e 5.2 a seguir:

$$Precision = \frac{\#VP}{\#VP + \#FP} \quad (5.1)$$

$$Recall = \frac{\#VP}{\#VP + \#FN} \quad (5.2)$$

Ainda para a identificação do que é um verdadeiro positivo e um falso negativo, sem a real avaliação de um usuário, é possível calcular através do conceito de *Ground truth* (Verdade de Campo). *Ground truth* é um termo usado em vários campos para se referir a informações fornecidas por observação direta ao contrário da informação fornecida por inferência. Também é usado em modelos estatísticos para provar ou refutar hipóteses de investigação (BUCKLAND; GEY, 1994). Assim, para definição de *Ground truth* dos resultados dos algoritmos de recomendação com base em distância, média de avaliação e tempo médio de atendimento. Para cada um foi definido um critério de *Ground truth* conforme demonstrado a seguir:

- **Distância:**

- $\#VP$: Quantidade de elementos que o algoritmo recomendou e que a distância é menor que a aceitável.
- $\#FP$: Quantidade de elementos que o algoritmo recomendou e que a distância é maior que a aceitável.

- $\#FN$: Quantidade de elementos que não o algoritmo não recomendou, mas que a distância é menor que a aceitável.

- **Tempo de Atendimento:**

- $\#VP$: Quantidade de elementos em que o algoritmo recomendou e que o tempo de atendimento é menor que a média do dia.
- $\#FP$: Quantidade de elementos em que o algoritmo recomendou e que o tempo de atendimento é maior que a média do dia.
- $\#FN$: Quantidade de elementos em que o algoritmo não recomendou e que o tempo de atendimento é menor que a média do dia.

- **Avaliação:**

- $\#VP$: Quantidade de elementos em que o algoritmo recomendou e que a avaliação é melhor que a média do dia.
- $\#FP$: Quantidade de elementos em que o algoritmo recomendou e que a avaliação é abaixo da média do dia.
- $\#FN$: Quantidade de elementos em que o algoritmo não recomendou e que a avaliação é acima da média do dia.

Segundo (GUNAWARDANA; SHANI, 2009) geralmente, é esperado um *trade off*¹ entre essas quantidades. Enquanto permitir longas listas de recomendação melhora o *Recall*, isto também é uma medida que reduz o *Precision*.

5.3 Implementação

O objetivo principal de avaliação é saber se o método de recomendação utilizado no *CrowdHealth* funciona e como é o seu desempenho. Portanto, foi implementado a parte de recomendação do modelo *CrowdHealth*, que é a parte principal do sistema. Motivar multidões a avaliar um sistema é uma tarefa difícil e uma boa avaliação depende de uma amostra de quantidade significativa. Por este motivo, optou-se por utilizar simulação na avaliação do *CrowdHealth*. Utilizando-se de uma fonte de dados já existente. Esta mesma fonte foi empregada no trabalho de Yang (YANG et al., 2015). Os dados estão num arquivo no formato *tsv*², como apresentado abaixo.

1. ID do usuário - anonimizado
2. ID do local - Foursquare

¹expressão do inglês que significa relação custo/benefício

²o formato *tsv* é semelhante ao *csv*, porém as informações são separadas por tabulações.

3. ID do local - Foursquare (Essa informação aparece repetida)
4. Nome da categoria do local - Fousquare
5. Latitude
6. Longitude
7. Compensação do Fusorário em minutos (A compensação em minutos entre quando o *checkin* ocorreu em um mesmo tempo no UTC)
8. UTC³ tempo
9. Intervalo
10. Avaliação

Desta forma, simulou-se a requisição de recomendações de usuários usando um *dataset* real contendo informações do *Foursquare*⁴. O arquivo *dataset_TSMC2014_NYC.txt* possuía 227428 *check-ins*. Esses registros foram obtidos através de uma coleta de 10 meses na cidade de Nova Iorque, EUA durante o período de 12 de abril de 2012 até 16 de fevereiro de 2013 (YANG et al., 2015). O arquivo foi dividido em duas partes: A primeira parte representava os *check-ins* realizados pelos usuários nos centros clínicos. Ao passo que a segunda parte, representava usuários requisitando por recomendações de centros clínicos em outros locais. A informação que identifica esses registros está na quarta tabulação (4. *Nome da categoria do local Fousquare*) do arquivo. Quando esta informação possui *Medical Center*, conforme apresentado abaixo, então considerou-se como sendo um registro de usuários já localizados em centros clínicos, recebendo atendimento. E quando não possuíam, considerou-se como sendo usuários, em outros locais, buscando por recomendações de centros clínicos. Assim, foram criadas funções para simular os processos assíncronos que acontecem no componente *Offline Recommender Service*, para cálculo do tempo de atendimento e avaliação dos centros clínicos. E também simulou-se usuários requisitando por recomendações de centros clínicos em outros locais, que acontecem no componente *Online Recommender Service*.

1	1541
2	4f0fd5a8e4b03856eeb6c8cb
3	4bf58dd8d48988d10c951735
4	Medical Center
5	35.705101088587135
6	139.6195900440216
7	540
8	Tue Apr 03 18:17:18
9	+0000
10	2012

³Tempo Universal Coordenado, do inglês Universal Time Coordinated

⁴<https://pt.foursquare.com/>

Função abaixo, implementada em *Python* realiza o cálculo da distância entre dois pontos, latitude e longitude. A distância está sendo medida de forma absoluta e não através da distância do caminho percorrido. Embora isso possa ser realizado utilizando-se a *API Google Maps*⁵ isso ficou como trabalho futuro visto que a ideia aqui é avaliar o método de recomendação e não sua implementação, apesar de que isso possa ser realizado utilizando-se da API do Google. A função de Bárcara, demonstrada abaixo, recebe os parâmetros de latitude e longitude de dois pontos e aplica a fórmula de Bárcara. Ela devolve o resultado em graus e, por isso, é multiplicado pela constante 1852 para realizar a conversão de graus para metros. A função `abs`, do *Python* exibida na linha 6, devolve o valor absoluto de um número.

```

1 def distance(locU, locC):
2     a = locU.x - locC.x
3     b = locU.y - locC.y
4     #1852 * para converter de graus pra metros
5     #6371 = 5KM
6     return (abs(math.hypot(a, b))/6371*abs(math.hypot(a, b)))*1852

```

A função abaixo, implementada na linguagem de programação do *Oracle, PLSQL*, representa o algoritmo de cálculo do tempo médio de atendimento. Ela calcula para cada centro clínico o tempo de atendimento que um usuário ficou naquele local. A função `AVG`, exibida na linha 10, do *PLSQL* devolve a média aritmética de um conjunto de valores.

```

1 create or replace function avgTimeService(pVENUEID ACTION_NY.VENUEID%type ,
2                                           pDtReq ACTION_NY.UTCTIME2%type ,
3                                           pdevFactor float ,
4                                           pUserWeight float) return number
5 as
6 r number;
7 begin
8
9     select
10          AVG(((act.utctimeoff - act.utctime2))*
11              (pDtReq - act.utctime2)*
12              pdevFactor *
13              pUserWeight
14          )
15          into
16          r
17 from
18     ACTION_NY act
19 where
20     VENUECATEGORYNAME like 'Medical_Center' and
21     VENUEID like pVENUEID
22 group by
23     VENUEID;

```

⁵<https://developers.google.com/maps/documentation/distance-matrix/?hl=pt-br>

```

24
25     return r*-1;
26
27 end;

```

Retorna a média ponderada das avaliações feitas para o local. O que pondera as avaliações é o quão atual é uma informação. É levado em consideração o quanto essa informação é interessante para o usuário.

```

1  create or replace function avgRating (pVENUEID ACTION_NY.VENUEID%type ,
2      pDtReq ACTION_NY.UTCTIME2%type ,
3      pdevFactor float ,
4      pUserWeight float) return number
5  as
6  r number;
7  begin
8      select
9          avg( act.rating *
10             (pDtReq - act.utctime2)*pdevFactor)*pUserWeight
11      into
12          r
13      from CENTROS_NY act
14      where VENUEID like pVENUEID
15      group by VENUEID;
16
17      return r;
18
19 end;

```

5.4 Pesquisa as Cegas

O cálculo da pesquisa as cegas realizado da seguinte forma. Cada centro clínico recebeu um número em ordem crescente. Gerou-se um número aleatório entre 1 e o número de possibilidades de centros clínicos a escolher para cada requisição. No caso, existem 868 centros clínicos de saúde distintos no *dataset*. Depois, foi aplicado os conceito de Verdadeiros e Falsos Positivos, bem como verdadeiros e falsos negativos para obtenção do *Precision* e *Recall*.

5.5 Arquitetura dos Experimentos

A arquitetura dos testes é dividida em duas partes, *hardware* e *software*. Na parte de *hardware*, as simulações foram executadas em uma máquina com processador *Core i3-3217U CPU @1.8GHz* de dois núcleos, sistema *Windows 10 64-bits*, 4GB de memória RAM e e 500GB de espaço em disco. Na segunda parte foi utilizado os seguintes *softwares* para implementação: Linguagem *Python* versão 2.7.12, o banco de dados *Oracle* versão *11g Express Edition* e o

middleware cx_Oracle versão 5.0.2, para a conexão entre os dois. Foi utilizado também uma planilha eletrônica de cálculo para composição, comparação e geração dos gráficos e tabelas de resultados apresentados na seção seguinte.

5.6 Resultados

No total foram realizadas as simulações de 204685 requisições de recomendação para de centros clínicos por usuários, para cada um dos tipos de avaliação. Este valor corresponde a todos os registros que não representavam *check-ins* de usuários em centros clínicos. A estratégia de dividir o processamento das recomendações de acordo com a sincronia das informações foi importante. O algoritmo de recomendação de centros clínicos usando os três elementos disponíveis foi o mais demorado para execução, levando 983 minutos para execução. Destaca-se também que como não é objetivo do trabalho avaliar o desempenho dos algoritmos, as tabelas possuíam somente chave primária, sem nenhum índice que pudesse aumentar o desempenho. A Tabela 4 apresenta os resultados obtidos de forma sumarizada para as quantidades de Verdadeiros Positivos (#VP), Falsos Positivos (#FP), Falsos Negativos (#FN) que foram necessárias para ao cálculo das métricas *Precision* e *Recall*, também mostrado na Tabela 4.

Tabela 4: Resultados Absolutos

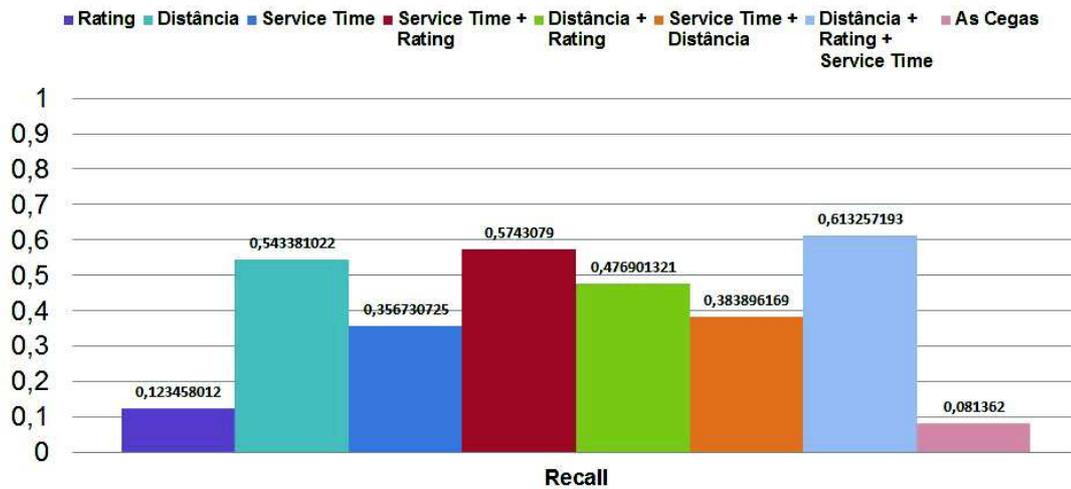
Crítérios de Avaliação / Métricas	#VP	#FP	#FN	Precision	Recall
Rating	18705	185979	132804	0,091384769	0,123458012
Distance	107753	96931	90548	0,526435872	0,543381022
Service Time	38518	166167	69457	0,18818184	0,356730725
Service Time + Rating	121442	83242	110016	0,593314573	0,524682664
Distance + Rating	63064	141620	69173	0,3081042	0,476901321
Service Time + Distance	76165	128519	122235	0,37211018	0,383896169
Distance + Rating + Service Time	117700	86984	74226	0,575032733	0,613257193
As Cegas	423	11190	394	0,025132	0,081362

Fonte: Elaborado pelo autor.

As Figuras 17 e 18 exibem os resultados de *Precision* e *Recall* de forma comparativa. É possível observar que o melhor método de recomendação foi também o método mais completo. A combinação dos três métodos Distância, Avaliação e Tempo de Atendimento (*Distance + Rating + Service Time*) 57,5% de *Precision*.

É possível perceber na 18 que o melhor método de recomendação também foi também o método mais completo para o *Recall*. A combinação dos três métodos Distância, Avaliação e Tempo de Atendimento (*Distance + Rating + Service Time*), alcançou até 61,33% de *Recall*. Este valor é aproximadamente 55% melhor que o pior método (*as Cegas*) e 3,45% melhor que o segundo melhor método (*Service Time + Rating*) para o *Precision*. Este método é também até 53% melhor que o pior método (*As Cegas*) e até 6,98% melhor que o segundo melhor método (*Distance*) para o *Recall*. O pior método, como já era esperado, foi a pesquisa as

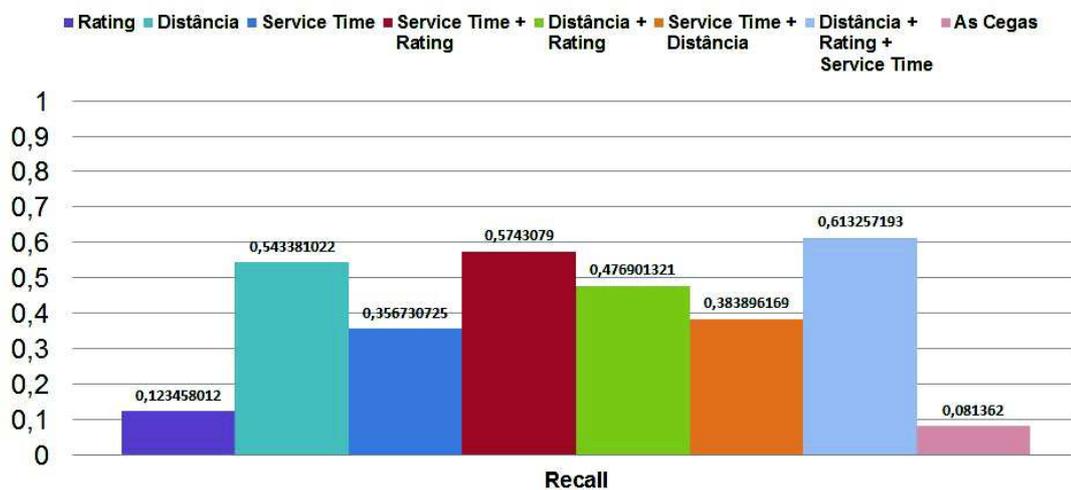
Figura 17: Comparativo entre Resultados - Precision



Fonte: Elaborado pelo autor.

cegas alcançando valores de 2,51% de *Precision* e 0,08% de *Recall*. Mas descontando essa avaliação, o segundo pior método é o de somente pela avaliação (*Rating*). Desta forma, o método mais completo *Distance + Rating + Service Time* apresentou os maiores valores de *Precision* e *Recall*. Portanto, isso significa que o algoritmo de recomendação de centros clínicos utilizado no *CrowdHealth* tem capacidade de recomendar itens relevantes para o usuário.

Figura 18: Comparativo entre Resultados - Recall



Fonte: Elaborado pelo autor.

6 CONSIDERAÇÕES FINAIS

Neste capítulo, serão apresentadas as conclusões deste trabalho, contribuições esperadas, assim como, algumas ideias que não foram implementadas mas que poderão servir de fonte de inspiração para trabalhos futuros de outros pesquisadores.

6.1 Conclusões

A principal contribuição do modelo *CrowdHealth* é possibilitar a recomendação de centros clínicos de saúde baseado na combinação de múltiplos contextos, incluindo a localização, o tempo de atendimento e a avaliação dos centros clínicos de saúde. Isso é expresso de forma científica através das fórmulas de cálculo de recomendação de centros clínicos de acordo com a avaliação, tempo de espera e distância. O modelo traz a possibilidade de criar uma vantagem no compartilhamento das informações pessoais providas pelos usuários.

A Tabela 5 apresenta a comparação entre o *CrowdHealth* e os trabalhos mais significativos. Os trabalhos relacionados apresentados não exploram a questão de utilizar dados de fontes *crowdsourcing* e, em troca, propor diminuição no tempo de espera de serviços de saúde para seus usuários. Também não foi encontrado nenhum trabalho que faça uma recomendação de centros clínicos. É possível observar também que os PHR's estão presentes na maioria das publicações selecionados e que o *CrowdHealth* seria melhor aproveitado se fosse integrado a uma solução PHR.

CrowdHealth usa dos dados fornecidos pelos usuários para criar uma relação ganha-ganha. Os usuários oferecem as informações de tempo de atendimento e, com isso, sabem quais centros clínicos são os mais recomendados para eles. As recomendações acontecem usando três elementos, a localização do usuário, o tempo de atendimento informado pelo usuário e as avaliações que os usuários fazem dos atendimentos que receberam nos centros clínicos. A combinação dos três métodos Distância, Avaliação e Tempo de Atendimento (*Distance + Rating + Service Time*), alcançou até 61,33% de *Recall* e 57,5% de *Precision*, o que demonstra que o mecanismo de recomendação do *CrowdHealth* possui capacidade de recomendar itens interessantes para o usuário. Entretanto, o sucesso de um sistema de recomendação não depende unicamente da qualidade do mecanismo de recomendação. Alguns sistemas procuram modificar o comportamento do usuário o qual é influenciado por diversos outros parâmetros, o mais notável, é a interface do usuário.

6.2 Trabalhos Futuros

Durante este trabalho foram levantadas diversas possibilidades de abordagens que não puderam ser implementadas ou testadas. Algumas ideias são descritas a seguir, porém as possibilidades de estudos certamente não se esgotam nestas observações.

Tabela 5: CrowdHealth vs Trabalhos Relacionados

Sistema	CrowdHealth	Demorô	HealthGem	N/A	PeopleSave	Smart Food	N/A	FIHUO	CrowdHelp
Ano	2016	2015	2015	2015	2016	2014	2015	2016	2014
Autor Principal	R. Pereira	Oliveira	Anthony	Song	Majethia	Moorhead	Papageorgiou	Sheng	Besaleva
Crowdsourcing	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim
Data Mining	Sim*	N/A	Classificação Textual (N-Gram)	N/A	Classificação Textual (N-Gram)	N/A	N/A	Similaridade do Cosseno & self-organized map	N/A
Foco	Cidadãos e Agentes Públicos*	Cidadãos	Paciente	Pacientes e médicos	Pacientes e médicos	Pacientes e Agentes de saúde	Cidadãos	Cidadãos	Cidadãos
Multiplataforma	Sim*	Sim	Não	Não	Não	Sim	Sim	Sim	Sim
Objetivo	Mapear tempo de espera de centros clínicos de saúde de uma cidade e recomendar os mais adequados de acordo com o contexto do usuário	Mapear tempo de espera de serviços de uma cidade	Melhoria de hábitos de saúde através de compartilhamento de informações.	Construir um histórico médico pessoal de longo prazo para os pacientes.	Análise de revisão médica crowdsourcing para a recomendação e interrupção do uso de drogas.	Controle e gerenciamento de dietas alimentares.	Gerenciamento do estresse em cidades inteligentes.	Análise e comparação de dados relacionados a saúde	Assistência em tempo real para catástrofes como terremotos, furacões e outros desastre
PHR	Sim*	N/A	Sim	Sim	Sim	Sim	Sim	Sim	Não
Padrão de Cuidados em Saúde	2.0	N/A	1.0	0.0	0.0	2.0	1.0	1.0	1.0
Wait Time	Sim	Sim	Não	Não	Não	Não	Não	Não	Não

Fonte: Elaborado pelo autor.

- **m-PHR** - A integração das recomendações de centros clínicos seria melhor aproveitada se fosse incorporada a um sistema *m-PHR*. Uma possibilidade é o *Glucosio*¹ que é *open-source*.
- **Datamining** - Poderiam ser feitas diversas abordagens de *datamining* para cruzar informações dos exames dos PHRs com informações geográficas, tais como identificação de grupos de risco, identificação de padrões de doenças e grupos de risco, contextualizadas por região. Ou ainda, identificação qualitativa da saúde, respondendo a seguinte questão: "de onde pessoas com problemas de saúde semelhante ao meu foram melhor/pior atendidas?"
- **Outras abordagens de teste** - Avaliar os mesmos algoritmos de recomendação utilizando outras métricas de teste tais como Erro Quadrático Médio, Desvio Médio da Raiz Quadrada, *Click-Through Rate*.
- **Filtros Colaborativos** - Implementar a parte de filtros colaborativos, através de similaridades entre usuários e/ou centros clínicos.
- **Ludificação ou gametization** - Criar mecanismos de recompensa (ver Figura 1) para motivar ainda mais utilização do *CrowdHealth*.

¹<http://www.glucosio.org/>

- **Atendimento e Tempo de Espera** simulação com diferentes valores de desvalorização referente ao atendimento e tempo de espera.

6.3 Contribuições

CrowdHealth é um modelo para recomendação de centros clínicos para usuários de acordo com sua geolocalização, a avaliação de outros usuários e tempo de atendimento informado por outros usuários. A recomendação de centros clínicos através do *CrowdHealth* motiva seus usuários a utilizar e compartilhar as informações de seus PHR's, a fim de contribuir com a saúde pública através de uma abordagem de *crowdsourcing*. A motivação dos usuários vem através do auxílio na tomada de decisão acerca de qual centro clínico é o mais adequado para ele.

O modelo do *CrowdHealth* é disponibilizado como um serviço para atender pacientes e cidadãos no geral. Dentre outros ganhos é possível destacar:

- Os pacientes ganham ao saber qual unidade de saúde mais próxima é mais adequada a sua necessidade. Adequado pode ser tanto um atendimento mais rápido, quanto um atendimento mais perto, ou ainda de qualidade superior conforme relatos de outros usuários.
- Os profissionais de saúde ganham ao saber em quais aspectos impactam mais na qualidade do serviço oferecido aos pacientes, ou ainda, quais são os aspectos a melhorar que a população mais reclama. Também terão acesso a informações quantitativas de dados de saúde dos usuários.
- Órgãos públicos ganham ao ter acesso a informações precisas sobre a qualidade do serviço de saúde oferecido a população, descobrimento e identificação de padrões de doenças e grupos de risco, contextualizadas por região.
- Por fim, o serviço de saúde oferecido aos cidadãos também recebe um ganho de qualidade, ao promover um balanceamento da procura de seus usuários que procurarão centros de saúde disponíveis conforme suas necessidades.

Também existem ganhos diretos tais como: diminuição do tempo de espera de atendimento; Balanceamento na utilização dos centros clínicos da cidade e conseqüentemente diminuição do tempo de espera de atendimento. E indiretos como: aumento na qualidade da saúde da cidade e aceleração da taxa de adoção dos *PHRs*.

REFERÊNCIAS

- ADIDA, B.; SANYAL, A.; ZABAK, S.; KOHANE, I. S.; MANDL, K. D. Indivo x: developing a fully substitutable personally controlled health record platform. **AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium**, v. 2010, p. 6–10, 2010.
- AGRAWAL, D.; DAS, S.; El Abbadi, A. Big data and cloud computing: current state and future opportunities. ... **of the 14th International Conference on ...**, p. 530–533, 2011.
- ANTELIO, M.; ESTEVES, M. G. P.; SCHNEIDER, D.; SOUZA, J. M. D. Qualitocracy: a data quality collaborative framework applied to citizen science. **Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics**, p. 931–936, 2012.
- ANTHONY, M.; PHILLIP, M.; MAE, C. THE USE OF SOCIAL NETWORKS IN AID OF PERSONAL HEALTH MONITORING. **DLSU Research Congress 2015**, v. 3, p. 1–6, 2015.
- BESALEVA, L. I.; WEAVER, A. C. CrowdHelp: m-health application for emergency response improvement through crowdsourced and sensor-detected information. **Wireless Telecommunications Symposium**, 2014.
- BRABHAM, D. C. **Crowdsourcing**.]: Mit Press, 2013.
- BUCKLAND, M.; GEY, F. The Relationship Between Recall and Precision. **J. Am. Soc. Inf. Sci.**, New York, NY, USA, v. 45, n. 1, p. 12–19, Jan. 1994.
- CARAGLIU, A.; DEL BO, C.; NIJKAMP, P. Smart cities in Europe. **Journal of urban technology**, v. 18, n. 2, p. 65–82, 2011.
- CAVNAR, W. B.; TRENKLE, J. M. et al. N-gram-based text categorization. **Ann Arbor MI**, v. 48113, n. 2, p. 161–175, 1994.
- CERRATO, P. **Why Personal Health Records Have Flopped**. 2012.
- CHATZIMILIOUDIS, G.; KONSTANTINIDIS, A.; LAOUDIAS, C.; ZEINALIPOUR-YAZTI, D. Crowdsourcing with smartphones. **IEEE Internet Computing**, v. 16, n. 5, p. 36–44, 2012.
- CHEN, C. P.; ZHANG, C.-Y. Data-intensive applications, challenges, techniques and technologies: a survey on big data. **Information Sciences**, University of Macau, Macau, China, v. 275, p. 314–347, 2014.
- CHEN, C.; WANG, Y. SPARC: strategy-proof double auction for mobile participatory sensing. **Proceedings - 2013 International Conference on Cloud Computing and Big Data, CLOUDCOM-ASIA 2013**, p. 133–140, 2013.
- CHOWDHURY, G. **Introduction to modern information retrieval**.]: Facet publishing, 2010.
- DEBNATH, S.; GANGULY, N.; MITRA, P. Feature weighting in content based recommendation system using social network analysis. **Proceeding of the 17th international conference on World Wide Web**, p. 1041–1042, 2008.

- DEYO, R. A.; CHERKIN, D. C.; CIOL, M. A. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. **Journal of clinical epidemiology**, v. 45, n. 6, p. 613–619, 1992.
- DO, N. V.; BARNHILL, R.; HEERMANN-DO, K. A.; SALZMAN, K. L.; GIMBEL, R. W. The military health system's personal health record pilot with Microsoft HealthVault and Google Health. **Journal of the American Medical Informatics Association**, v. 18, n. 2, p. 118–124, 2011.
- DONNELLY, K. SNOMED-CT: the advanced terminology and coding system for ehealth. **Studies in health technology and informatics**, v. 121, p. 279, 2006.
- EL FADLY, A.; DANIEL, C.; BOUSQUET, C.; DART, T.; LASTIC, P.-Y.; DEGOULET, P. Electronic Healthcare Record and clinical research in cardiovascular radiology. HL7 CDA and CDISC ODM interoperability. In: AMIA, 2007. **Anais...** 2007.
- ELMANGOUSH, A.; COSKUN, H.; WAHLE, S.; MAGEDANZ, T. Design aspects for a reference M2M communication platform for Smart Cities. In: INNOVATIONS IN INFORMATION TECHNOLOGY (IIT), 2013 9TH INTERNATIONAL CONFERENCE ON, 2013. **Anais...** 2013. p. 204–209.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996.
- GOMES, C.; SCHNEIDER, D.; De Souza, J.; XEXEO, G. Cassino musical: a game with a purpose for social recruitment and measurement of musical talent. **Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2013**, p. 593–598, 2013.
- GOMES, C.; SCHNEIDER, D.; De Souza, J.; XEXÉO, G. Towards a framework for crowdsourced collection, cleaning and measurement of digital content. **Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013**, p. 1032–1039, 2013.
- GOMES, C.; SCHNEIDER, D.; MORAES, K.; De Souza, J. Crowdsourcing for music: survey and taxonomy. **Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics**, p. 832–839, 2012.
- GUNAWARDANA, A.; SHANI, G. A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. **The Journal of Machine Learning Research**, v. 10, p. 2935–2962, 2009.
- GUNTER, T. D.; TERRY, N. P. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. **Journal of Medical Internet Research**, v. 7, n. 1, 2005.
- LEE, L.-H.; CHOU, Y.-T.; HUANG, E.-W.; LIOU, D.-M. Design of a personal health record and health knowledge sharing system using IHE-XDS and OWL. **Journal of medical systems**, v. 37, n. 2, p. 1–12, 2013.
- LUO, X.; OHYAMA, W.; WAKABAYASHI, T.; KIMURA, F. Automatic Chinese Text Classification Using Character-Based and Word-Based Approach. In: INTERNATIONAL

CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION, 2013., 2013.
Anais... 2013. p. 329–333.

MAJETHIA, R. PeopleSave : recommending effective drugs through web crowdsourcing.
COMSNETS 2016 - NetHealth Workshop, p. 1–6, 2016.

MCDONALD, C. J.; HUFF, S. M.; SUICO, J. G.; HILL, G.; LEAVELLE, D.; ALLER, R.;
FORREY, A.; MERCER, K.; DEMOOR, G.; HOOK, J. et al. LOINC, a universal standard for
identifying laboratory observations: a 5-year update. **Clinical chemistry**, v. 49, n. 4,
p. 624–633, 2003.

MEDEIROS, S. P. J.; SCHNEIDER, D.; De Souza, J. M.; ESTEVES, M. G. P. Strategic
planning in semantic crowdware large groups decision. **Proceedings of the 2012 IEEE 16th
International Conference on Computer Supported Cooperative Work in Design,
CSCWD 2012**, p. 649–654, 2012.

Microsoft Corporation. **Microsoft Application Architecture Guide**. 2nd Editio. ed. 2009.
560 p.

MOORHEAD, A.; BOND, R. Smart Food : crowdsourcing of experts in nutrition and
non-experts in identifying calories of meals using smartphone as a potential tool contributing
to obesity prevention and management. **Journal of Telemedicine and Telecare**, v. 16, n. 8,
p. 2–4, 2014.

MORAES, A. L. D.; FONSECA, F.; ESTEVES, M. G. P.; SCHNEIDER, D.; De Souza, J. M.
A meta-model for crowdsourcing platforms in Data Collection and Participatory Sensing.
**Proceedings of the 2014 IEEE 18th International Conference on Computer Supported
Cooperative Work in Design, CSCWD 2014**, p. 429–434, 2014.

NAPHADE, M.; BANAVAR, G.; HARRISON, C.; PARASZCZAK, J.; MORRIS, R. Smarter
cities and their innovation challenges. **Computer**, v. 44, n. 6, p. 32–39, 2011.

NASCIMENTO, P.; AGUAS, R.; SCHNEIDER, D.; De Souza, J. An approach to
requirements categorization using Kano's model and crowds. **Proceedings of the 2012 IEEE
16th International Conference on Computer Supported Cooperative Work in Design,
CSCWD 2012**, p. 387–392, 2012.

NASCIMENTO, P.; OSIEK, B.; XEXÉO, G. Análise de sentimento de tweets com foco em
notícias. **Revista Eletrônica de Sistemas de Informação ISSN 1677-3071 doi:
10.5329/RESI**, v. 14, n. 2, 2015.

OFFICE, C. B. Evidence on the Costs and Benefits of Health Information Technology.
Evidence report Technology Assesment, v. 132, n. May, p. 1–71, 2008.

OLIVEIRA, L. F.; SCHNEIDER, D.; De Souza, J. M.; RODRIGUES, S. A. Leveraging the
crowd collaboration to monitor the waiting time of day-to-day services. **Proceedings of the
2015 IEEE 19th International Conference on Computer Supported Cooperative Work in
Design, CSCWD 2015**, p. 109–114, 2015.

PAPAGEORGIU, A.; ZIGOMITROS, A.; PATSAKIS, C. Personalising and Crowdsourcing
Stress Management in Urban Environments via s-Health. **2015 6th International Conference
on Information, Intelligence, Systems and Applications (IISA)**, v. 3, p. 1–4, jul 2015.

PEREIRA, C. V.; ESTEVES, M. G. P.; MEDEIROS, S. P. J.; De Souza, J. M.; ANTELIO, M. How the crowd can change collaborative work in patient care. **Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2013**, n. 1, p. 527–532, 2013.

PEREIRA, J. L. ANÁLISE PREDITIVA EM SISTEMAS DE INFORMAÇÃO NO CONTEXTO DO BIG DATA. **CENTRO UNIVERSITARIO EURIDES DE MARIA**, 2015.

PÉREZ-MARTÍNEZ, P. A.; MARTÍNEZ-BALLESTÉ, A.; SOLANAS, A. Privacy in Smart Cities-A Case Study of Smart Public Parking. In: PECCS, 2013. **Anais...** 2013. p. 55–59.

POULYMENOPOULOU, M.; MALAMATENIOU, F.; PRENTZA, A.; VASSILACOPOULOS, G. Challenges of Evolving Pincloud Phr Into a Phr-Based Health Analytics System. **European, Mediterranean & Middle Eastern Conference on Information Systems 2015**, v. 2015, p. 1–8, 2015.

PRILLA, M.; RITTERSKAMP, C. Finding Synergies: web 2.0 and collaboration support systems. In: COOP, 2008. **Proceedings...** 2008. v. 8.

QUINN, A. J.; BEDERSON, B. B. Human computation. **Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11**, p. 1403, 2011.

RAJARAMAN, A.; ULLMAN, J. D. Mining of Massive Datasets. **Lecture Notes for Stanford CS345A Web Mining**, v. 67, p. 328, 2011.

ROUSE, M. **Crowdsourcing**. 2016.

SAWYER, A.; AYERS, S.; FIELD, A. P. Posttraumatic growth and adjustment among individuals with cancer or HIV/AIDS: a meta-analysis. **Clinical Psychology Review**, v. 30, n. 4, p. 436–447, 2010.

SCHNEIDER, D.; De Souza, J.; LUCAS, E. M. Towards a typology of social news apps from a Crowd Computing perspective. **Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics**, v. 2014-Janua, n. January, p. 1134–1140, 2014.

SCHNEIDER, D.; MORAES, K.; De Souza, J. M.; ESTEVES, M. G. P. CSCWD: five characters in search of crowds. **Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2012**, p. 634–641, 2012.

SHENG, M.; ZHANG, Y.; YANG, J.; LI, C.; XING, C. Fihuo: a mobile smart health service platform. **Proceedings - 2015 12th Web Information System and Application Conference, WISA 2015**, p. 179–183, 2016.

SOLANAS, A.; PATSAKIS, C.; CONTI, M.; VLACHOS, I.; RAMOS, V.; FALCONE, F.; POSTOLACHE, O.; PEREZ-MARTINEZ, P.; PIETRO, R.; PERREA, D.; MARTINEZ-BALLESTE, A. Smart health: a context-aware health paradigm within smart cities. **IEEE Communications Magazine**, v. 52, n. 8, p. 74–81, 2014.

SONG, Y.-T.; HONG, S.; PAK, J. Empowering patients using cloud based personal health record system. **2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). Proceedings**, p. 293 – 8, 2015.

YANG, D.; ZHANG, D.; ZHENG, V. W.; YU, Z. Modeling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs. **IEEE Transactions on Systems, Man, and Cybernetics: Systems**, v. 45, n. 1, p. 129–142, 2015.

ZHANG, H.; LAW, E.; MILLER, R. C.; GAJOS, K. Z.; PARKES, D. C.; HORVITZ, E. Human Computation Tasks with Global Constraints. **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**, p. 217–226, 2012.