

**CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA
PAULA SOUZA**

Faculdade de Tecnologia Rubens Lara

**Curso Superior de Tecnologia em Análise e
Desenvolvimento de Sistemas**

HUGO BESSA SILVA DE OLIVEIRA

**SISTEMA DE EDIÇÃO COLABORATIVA POR P2P
UTILIZANDO PARADIGMA DE PROGRAMAÇÃO
FUNCIONAL**

**Santos, SP
2016**

HUGO BESSA SILVA DE OLIVEIRA

**SISTEMA DE EDIÇÃO COLABORATIVA POR P2P
UTILIZANDO PARADIGMA DE PROGRAMAÇÃO
FUNCIONAL**

Pré-Projeto de Desenvolvimento apresentada à Faculdade de Tecnologia Rubens Lara, como exigência para a obtenção do Título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientador: Prof. Me. Alexandre Garcia de Oliveira

Santos, SP

2016

SUMÁRIO

1. INTRODUÇÃO	3
1.1. OBJETIVO GERAL	5
1.1.1.OBJETIVOS ESPECÍFICOS	5
1.2. ORGANIZAÇÃO.....	5
2. REVISÃO BIBLIOGRÁFICA.....	6
2.1. CONFLICT-FREE REPLICATED DATA TYPES.....	6
2.2. LOGOOT.....	8
2.3. ARQUITETURA DE REDE PEER-TO-PEER.....	9
2.4. PARADIGMA DE PROGRAMAÇÃO FUNCIONAL.....	10
3. METODOLOGIA.....	11
3.1. CASOS DE USO	11
3.1.1. DIAGRAMA DE CASO DE USO	11
4. REFERÊNCIAS BIBLIOGRÁFICAS	12

1. INTRODUÇÃO

A difusão de dispositivos móveis leva, ano após ano, o acesso à internet para cada vez mais pessoas. "Em 2015 existem mais de 7 bilhões de assinaturas de celulares móveis em todo o mundo, partindo de menos de 1 bilhão em 2000." (SANOU; BRAHIMA, 2015, p. 1, tradução nossa).

Pela primeira vez, mais de 2 bilhões de pessoas tiveram acesso à rede mundial de computadores. Isso torna possível a utilização de *softwares online*, como clientes de *e-mail*, sistemas de gerenciamento de conteúdo e sistemas de edição colaborativa de documentos. Disponibilizadas na internet, estas aplicações podem ser utilizadas em diferentes tipos de dispositivos e conectar usuários distantes fisicamente.

Algumas aplicações distribuídas utilizam o modelo de computação cliente-servidor, também conhecida como cliente-servidor de três níveis. Nesta arquitetura, todos os clientes, solicitantes de um recurso, se conectam e se comunicam através de um servidor, fornecedor de recursos.

Segundo Donald Bales (2003, p. 133, tradução nossa) "Uma arquitetura de três níveis adiciona ao modelo de dois níveis uma nova camada que isola o processamento de dados em um lugar centralizado [...]". Este isolamento traz importantes vantagens para aplicações distribuídas. Entretanto, este modelo computacional apresenta problemas graves quando utilizado em redes de internet móveis.

Redes de internet móveis são geralmente mais lentas do que redes banda-larga cabeadas. Além disso, a cobertura geográfica e a qualidade da conexão variam drasticamente entre regiões — incluindo locais não muito distantes. Dados da OpenSignal (2016) mostram uma latência média de 172 milissegundos em redes 3G. O modelo de comunicação cliente-servidor fica, então, prejudicado: com a indisponibilidade da rede e grande latência para comunicação, a experiência em aplicações que dependem da interação constante com o servidor é degradada.

De acordo com Shapiro et al. (2011) modelos de Forte Consistência Eventual (FCE) garantem melhor disponibilidade e performance em sistemas distribuídos, principalmente em redes com grande latência. Isto é alcançado por atualizações

serem realizadas em réplicas locais dos dados e eventualmente sincronizadas com outras réplicas dentro do sistema. Uma estrutura de dados compatível com as exigências de FCE é a *Conflict-free Replicated Data Type* (CRDT, em português Tipo de Dados Replicados Livre de Conflitos).

CRDTs garantem a consistência dos dados pois operações realizadas nesta estrutura são comutativas¹. CRDTs baseadas em operações transmitem seu estado propagando operações de atualização realizadas. Estas operações também devem ser comutativas. Para remover a necessidade de manter garantias de entrega única das mensagens de operação, as funções de operações de atualização também devem ser idempotentes² e associativas³.

Com as garantias de uma CRDT com operações associativas, comutativas e idempotentes, é possível construir um sistema em uma arquitetura de rede em que as réplicas se comunicam diretamente e propagam atualizações de forma distribuída e assíncrona. Esta arquitetura de rede se chama *peer-to-peer* (P2P), onde cada participante da rede atua tanto como cliente (requisitando dados) quanto como servidor (servindo dados) (SCHOLLMEIER; Rüdiger, 2002).

O presente trabalho utiliza a CRDT Logoot para implementar um Sistema de Edição Colaborativa distribuído com comunicação por uma rede *peer-to-peer*. A Logoot utiliza simples identificadores de posição para atingir comutatividade e associatividade em operações de inserção e remoção, que podem ser executadas em qualquer réplica participante do sistema sem alterar a ordem das linhas do documento (WEISS; URSO; MOLLI, 2008, p. 5).

Com o fim de facilitar a implementação da estrutura de dados Logoot, da interface de rede e do editor de texto necessários para o Sistema de Edição Colaborativa a ser desenvolvido como estudo de caso deste projeto, Elm foi a escolhida. Elm é uma linguagem de programação fortemente tipada que segue o paradigma funcional, e o projeto tira proveito de suas estruturas de dados imutáveis e tipos.

¹ "Dois elementos x e y de um conjunto S são comutativos sob uma operação $*$ se eles satisfazem $x * y = y * x$ " (WEISSTEIN; ERIC, tradução nossa)

² Uma operação é idempotente quando a sua aplicação repetida tem o mesmo efeito que uma (FOLDOC, 1995, tradução nossa)

³ "[...] associativa: $a(bc) = (ab)c$ para todo $a, b, c \in G$ " (HUGENFORD; Thomas, 1974, p. 24, tradução nossa)

1.1. OBJETIVO GERAL

Desenvolver um sistema de edição colaborativa Logoot em um editor de texto com comunicação *peer-to-peer* (P2P) utilizando a linguagem de programação Elm.

1.1.1.OBJETIVOS ESPECÍFICOS

Desenvolver módulo Elm que implemente o sistema Logoot e permita todas as suas funcionalidades: adição, remoção e edição de conteúdo; Implementar módulo Elm que transforme a estrutura de dados Logoot em um editor de texto capaz de realizar operações de adição, remoção e edição de conteúdo no sistema; Implementar infraestrutura de rede pra suportar a Edição Colaborativa por P2P utilizando a tecnologia WebRTC; Demonstrar a utilização de uma linguagem de programação funcional para trabalhar com estruturas de dados complexas e compor pequenas partes do *software* em um sistema completo.

1.2. ORGANIZAÇÃO

A introdução do tema do trabalho está no primeiro capítulo. Nela temos a apresentação do tema do trabalho e seus objetivos.

No segundo capítulo a revisão bibliográfica é apresentada — um detalhamento de cada tecnologia utilizada para solucionar o problema apresentado. Estão neste capítulo assuntos como *Conflitc-free Replicated Data Types*, arquitetura de rede *peer-to-peer*, Logoot e paradigma de programação funcional.

2. REVISÃO BIBLIOGRÁFICA

2.1. CONFLICT-FREE REPLICATED DATA TYPES

Em sistemas distribuídos é muito custoso manter uma ordem global de operações realizadas em uma estrutura de dados. Modelos de Forte Consistência Eventual (FCE) são mais adequados em sistemas distribuídos que têm como requisito a replicação e consistência de dados, principalmente quando diferentes clientes estão conectados por meio de redes instáveis com grande latência (SHAPIRO et al., 2011).

Como descrito por Shapiro et al. (2011), um modelo de Forte Consistência Eventual estabelece a convergência livre de conflitos de operações distribuídas em uma estrutura de dados. Dessa forma, réplicas não necessitam de sincronização e o sistema se mantém consistente independente da ocorrência de inúmeras falhas. Na FCE, réplicas que receberam as mesmas atualizações eventualmente atingem o estado de convergência.

CRDTs são estrutura de dados que garantem a compatibilidade com o modelo de FCE. Nelas, operações podem ser realizadas localmente e eventualmente sincronizadas com outras réplicas, sempre mantendo a consistência entre as estruturas de dados distribuídas.

Segundo Weiss, Urso, Molli (2008 apud Sun et al. 1998 [adicionar referência], p. 68, tradução nossa):

Um [sic] sistema de edição colaborativa é considerado correto se este respeitar o critério CCI:

- **Causality** (causalidade, em português): todas as operações ordenadas por uma relação de precedência, no sentido da relação *aconteceu antes* de Lamport [adicionar referência], são executadas na mesma ordem em todas as réplicas.
- **Convergence** (convergência, em português): O sistema converge se todas as réplicas são idênticas quando o sistema está parado [...]
- **Intention preservation** (preservação de intenção, em português): O efeito esperado de uma operação deve ser observado em todas as réplicas [...]

Existem dois tipos principais de CRDTs: baseadas em estado e baseadas em operações.

CRDTs baseadas em estado propagam seu estado completo entre réplicas, as quais utilizam uma função de combinação para aplicar alterações. A função de

combinação deve ser associativa, comutativa⁴ e idempotente para assegurar as propriedades de convergência da FCE. Numa abordagem baseada em estado, clientes propagam diferentes estados de sua estrutura de dados, sem incluir informações sobre a evolução deste estado (AHMED-NACER et al., 2011).

Sendo c a função de combinação, s_x , s_y e s_z o estado de três réplicas x , y e z , então $c(s_x, s_y) = c(s_y, s_x)$ (comutatividade), $c(s_x, c(s_y, s_z)) = c(c(s_x, s_y), s_z)$ (associatividade) e $c(s_x, s_y) = c(s_x, s_y)$ (idempotência).

As CRDTs baseadas em operações propagam operações como inserção e remoção entre réplicas, em vez de todo o estado. A convergência é garantida por assegurar que a função de aplicação de operação é uma função associativa, comutativa e idempotente.

Sendo o a função de aplicação de uma operação, s o estado de uma réplica, i uma operação de inserção e r uma operação de remoção, então $o(o(s, i), r) = o(o(s, r), i)$ (comutatividade e associatividade) e $o(s, i) = o(s, i)$ (idempotência).

Escolher entre CRDTs baseadas em estado ou baseadas em operações exige um entendimento melhor sobre as vantagens e desvantagens de cada abordagem.

As CRDTs baseadas em estado possuem a vantagem de implementação de apenas uma função de união para convergir estados de duas réplicas. Como todo o estado é propagado, também não há a necessidade de manter um *buffer* com todas as atualizações realizadas no documento. Entretanto, em sistemas em tempo-real, a sobrecarga causada pela propagação de todo o estado a cada atualização pode deixar o sistema menos responsivo, já que a diferença entre os estados de cada réplica precisa ser calculado constantemente (AHMED-NACER et al., 2011).

Para minimizar esta sobrecarga, Almeida, Shoker e Baquero (2014) propõem a *Delta State-based CRDT* (δ -CRDT). Este tipo de dados permite que apenas diferenças (*deltas*, em inglês) sejam propagadas entre réplicas. As diferenças em δ -CRDTs são geradas pelos chamados *δ -mutators*, que retornam o resultado das mutações realizadas no estado.

CRDTs baseadas em operações não impõem grande sobrecarga na propagação de mensagens de atualização. Como CRDTs baseadas em operações armazenam cada operação durante a evolução de uma réplica, também não é necessário

⁴ De acordo com o critério de causalidade de sistemas de edição colaborativa, não é estritamente necessário que a função de combinação seja comutativa entre todas as versões do estado de um réplica. Se o critério de causalidade for respeitado, a função de combinação só precisará ser comutativa entre versões que não possuem uma relação de precedência.

realizar operações de combinação em toda a CRDT — apenas aplicar operações ainda não incorporadas (AHMED-NACER et al., 2011).

2.2. LOGOOT

Logoot é uma estrutura de dados criada para ser utilizada em sistemas de edição colaborativa em redes distribuídas *peer-to-peer*. Ela é também uma CRDT, tornando possível que várias cópias de um mesmo documento sejam alteradas e sincronizadas sem conflito (WEISS; URSO; MOLLI, 2008, p. 5, tradução nossa).

Uma vantagem da Logoot quando comparada com outras estruturas de dados como Wooki (WEISS; URSO; MOLLI, 2007) é que ela não adiciona um grande peso em cima dos dados brutos de um documento devido à baixa quantidade de metadados necessários para mantê-la consistente entre sincronizações com réplicas distribuídas. Isso deve-se à utilização de identificadores de posição (*pid*) baseados em uma lista de inteiros que podem ser removidos da CRDT sem afetar a ordem de outros identificadores remanescentes (WEISS; URSO; MOLLI, 2008, p. 5, tradução nossa).

A duas operações disponíveis em uma Logoot são *inserir* e *remover*. Para se adequar como uma CRDT, essas operações precisam ser idempotentes, comutativas e associativas. Uma implementação baseada no sistema proposto por Weiss, Urso e Molli (2008) não respeita a regra de comutatividade, levando réplicas a divergirem dependendo da ordem em que recebem operações realizadas por outras réplicas.

Dada as seguintes definições:

- $i = (pid, c)$: operação de inserção que insere um conteúdo c no identificador de posição pid ;
- $r = (pid)$: operação de remoção que remove o identificador de posição pid ;
- $apply(op, s)$: função que aplica a operação op no estado s e retorna o novo estado.
- s_i : estado inicial de uma réplica

Temos $apply(i(pid_1, ""), apply(r(pid_1), s_i)) \neq apply(r(pid_1), apply(i(pid_1, ""), s_i))$, já que as operações i e r não são comutativas. Isso se deve à implementação originalmente proposta ignorar a remoção de um pid ainda não existente no estado de uma réplica.

Na CRDT Logoot-Undo (WEISS; URSO; MOLLI, 2010) — uma evolução da Logoot que permite ações de *undo* (desfazer, em inglês) — os autores propõem a adição de relacionar cada *pid* com um *degree* (grau, em inglês). A adição do *degree* permite expressar na estrutura de dados quando um *pid* ainda não adicionado é removido. Dessa forma, quando uma réplica recebe uma operação remota de remoção de um *pid* e futuramente sua inserção, o efeito é o mesmo de quando ela receber a inserção primeiro.

Ao tornar as operações de inserção e remoção comutativas, é possível implementar um sistema de edição colaborativo sem garantir a causalidade destas operações — elas não são mais ordenadas por uma relação de precedência, já que podem ser executadas em qualquer ordem e produzir o mesmo resultado.

2.3. ARQUITETURA DE REDE *PEER-TO-PEER*

Definida por Schollmeier apud Kellerer, uma arquitetura de rede é *peer-to-peer* (P2P) se seus participantes agem tanto como clientes (requisitando recursos) quanto como servidores (fornecendo recursos), utilizando seus próprios recursos de hardware que são acessados diretamente por outros participantes da rede sem entidades intermediárias.

Este estilo de arquitetura ficou popularizada com o *software* de compartilhamento de arquivos de áudio Napster (SCHOLLMEIER, 2002, p. 1, tradução nossa).

"Eles [sistemas P2P] provêm escalabilidade de distribuição de conteúdo mais barata e resistência a tentativas de censura." (WEISS; URSO; MOLLI, 2010 apud ANDROUTSELLIS-THEOTOKIS; SPINELLIS, 2004 [adicionar referência]). No presente trabalho, esta arquitetura soluciona o problema da necessidade de um servidor centralizado para sincronizar réplicas, transformando cada usuário em uma réplica do sistema de edição colaborativa.

Sistemas de conteúdo dinâmico, como sistemas de edição colaborativa ou de controle de versionamento, ainda não são muito bem suportados em redes P2P (WEISS; URSO; MOLLI, 2010, p. 1), muito devido à dificuldade de implementação de sistemas que mantenham a consistência de dados entre réplicas. A CRDT Logoot foi introduzida por Weiss, Urso e Molli (2008) justamente para solucionar este caso de uso.

- adicionar sobre distribuir o aplicativo por P2P, como visto por IPFS.

2.4. PARADIGMA DE PROGRAMAÇÃO FUNCIONAL

Em linguagens de programação imperativas, programas são escritos para modificar o estado no seu contexto de execução. Logo, a execução de programas escritos com linguagens de programação imperativas é não determinístico. Um outro paradigma de programação propõem uma abordagem diferente: a aplicação de funções que transformam dados de forma determinística.

"[...] *linguagens funcionais* são linguagens declarativas em que o modelo básico de computação é a função" ([adicionar referência, HUDAK, tradução nossa]). O estado de um programa funcional é aplicado como argumento de funções. Linguagens de programação modernas possuem um conjunto de utilitários para facilitar a escrita de programas funcionais.

Uma propriedade muito importante das funções em linguagens funcionais, e outras linguagens de estilo puro e declarativo, é a transparência referencial. "O termo *transparência referencial* é frequentemente utilizado para descrever este estilo de programação [puro e declarativo] em que 'iguais podem ser trocados por iguais'." ([adicionar referência, HUDAK, tradução nossa]). Isso significa que a aplicação de uma função pode ser substituída pelo seu resultado sem alterar o comportamento do programa.

A transparência referencial é a essência de muitas das vantagens desta classe de linguagem de programação. A garantia de uma função sempre retornar o mesmo resultado quando aplicada com os mesmos argumentos, não dependendo do seu contexto de execução, é a chave para tornar programas funcionais altamente performáticos com o uso de computação paralela, por exemplo.

Tratar funções como valores de primeira-classe é bem comum entre linguagens funcionais. Com funções sendo tratadas como outros valores (como inteiros e *strings*), é possível criar funções que abstraem a aplicação de outras funções ou que retornam funções para serem utilizadas posteriormente. Este tipo de função que retorna e/ou recebe outras função se chama Função de Ordem Superior (*Higher order function*, em inglês). Abstrair e construir programas a partir deste simples tipo "[...] aumenta a modularidade, servindo como um mecanismo de juntar fragmentos

de um programa." ([adicionar referência, HUDAK, apud HUGHES, 1984, tradução nossa]).

Funções na linguagem de programação Elm ([adicionar referência]), por exemplo, utilizam o método de *currying*. Estas funções são sequências de funções que aplicam o valor de cada argumento retornando uma nova função para o argumento sucessor, até a exaustão da lista de argumentos. O *currying* permite a escrita concisa de programas que passam valores em uma cadeia de aplicação de funções, já que é relativamente fácil criar abstrações de outras funções.

3. METODOLOGIA

O presente trabalho foi desenvolvido utilizando a metodologia ágil de desenvolvimento de *software*. Segundo Beck et al. (2001), o desenvolvimento ágil valoriza "indivíduos e interações sobre processos e ferramentas, *software* funcional sobre documentação abrangente, colaboração de clientes sobre negociação de contratos, responder a mudanças sobre seguir um plano". Estes valores realçam a importância de *software* desenvolvido, testado e executado acima outros objetivos como a presença de uma documentação extensiva — esta que frequentemente não é atualizada frequentemente pelos participantes do projeto.

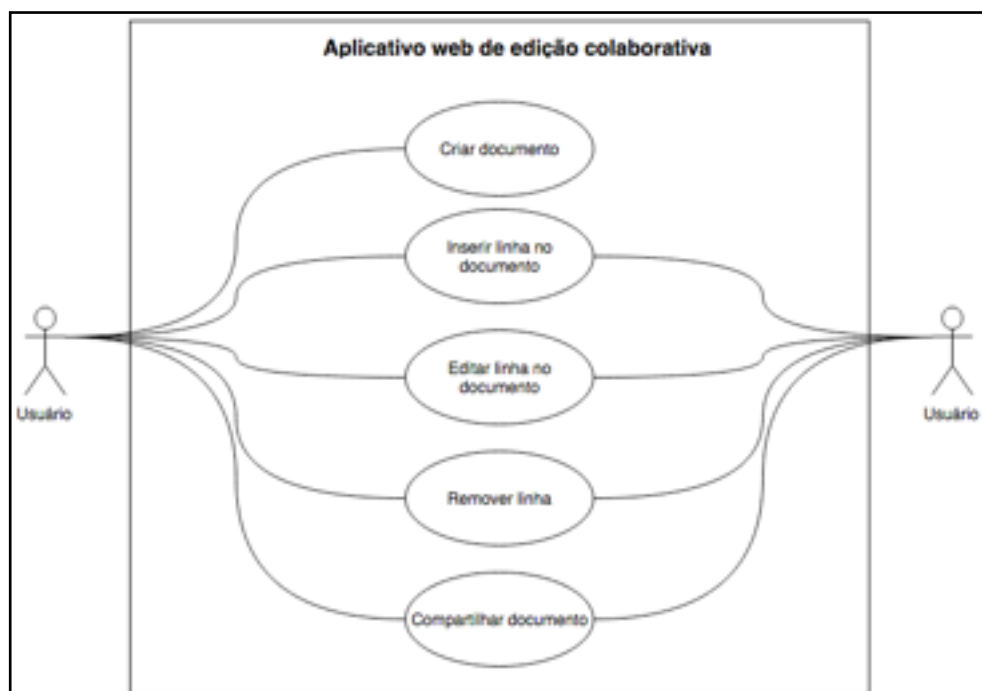
Projetos que adotam metodologias de desenvolvimento ágil também possuem documentação, embora possivelmente muito menos do que projetos que usam metodologias mais tradicionais baseadas em grandes quantidades de documentos e planejamento.

3.1. CASOS DE USO

3.1.1. DIAGRAMA DE CASO DE USO

Figura 1 - Diagrama de caso de uso

Fonte: OLIVEIRA, 2016



4. REFERÊNCIAS BIBLIOGRÁFICAS

ICT Data and Statistics Division. **ICT Facts & Figures 2015**. Maio de 2015. Disponível em: <<http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf>>. Acesso em: 10 de abril de 2016.

BALES, Donald. **JDBC Pocket Reference**. julho de 2003. Disponível em: <<https://web.archive.org/web/20061212141647/http://java.sun.com/developer/Books/jdbc/ch07.pdf>>. Acesso em: 29 de abril de 2016.

OpenSignal. **O Estado das Redes Móveis: Brasil (fevereiro 2016)**. fevereiro de 2016. Disponível em: <http://opensignal.com/reports/auto/data-2016-02-brazil/report_pt.pdf>. Acesso em: 16 de maio de 2016.

SHAPIRO, Marc; PREGUIÇA, Nuno; BAQUERO, Carlos; ZAWIRSKI, Marek. **Conflict-free Replicated Data Types**. 19 de julho de 2011. Disponível em: <<https://hal.inria.fr/inria-00609399v1/document>>. Acesso em: 10 de abril de 2016.

AHMED-NACER, Mehdi, IGNAT, Claudia-Lavinia, OSTER, Gérald, ROH, Hyun-Gul, URSO, Pascal. **Evaluating CRDTs for Real-time Document Editing**. Disponível em: <<https://hal.archives-ouvertes.fr/inria-00629503/document>>. Acesso em: 21 de setembro de 2016.

WEISSTEIN, Eric. **Commutative**. Disponível em: <<http://mathworld.wolfram.com/Commutative.html>>. Acesso em: 16 de abril de 2016.

FOLDOC. **Idempotent**. 11 de janeiro de 1995. Disponível em: <<http://foldoc.org/idempotent>>. Acesso em: 16 de abril de 2016.

HUNGERFORD, Thomas. **Algebra**. 1 ed. Cleveland: Springer, 1974.

SCHOLLMEIER, Rüdiger. **A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications**. 2002. Disponível em <<http://www.computer.org/csdl/proceedings/p2p/2001/1503/00/15030101.pdf>>. Acesso em: 10 de abril de 2016.

LETIA, Mihai; PREGUIÇA, Nuno; SHAPIRO, Marc. CRDTs: **Consistency without concurrency control**. junho de 2009. Disponível em: <<https://hal.inria.fr/inria-00397981/file/RR-6956.pdf>>. Acesso em: 1 de outubro de 2016.

WEISS, Stéphane; URSO, Pascal; MOLLI, Pascal. **Logoot: a P2P collaborative editing system**. 10 de dezembro de 2008. Disponível em: <<https://hal.inria.fr/inria-00336191/PDF/main.pdf>>. Acesso em: 11 de abril de 2016.

PHONESCOOP. **Smartphone**. Disponível em: <<http://www.phonescoop.com/glossary/term.php?gid=131>>. Acesso em: 16 de abril de 2016

WEISS, Stéphane; URSO, Pascal; MOLLI, Pascal. **Wooki: a P2P Wiki-based Collaborative Writing Tool**. 25 de junho de 2007. Disponível em: <<https://hal.inria.fr/inria-00156190v2/document>>. Acesso em: 13 de junho de 2016.

CLOJURE. **The Clojure Programming Language**. Disponível em: <<http://clojure.org/>>. Acesso em: 13 de junho de 2016.

BECK, Kent; BEEDLE, Mike; BENNEKUM, Arie van; COCKBURN, Alistair; CUNNINGHAM, Ward; FOWLER, Martin; GRENNING, James; HIGHSMITH, Jim; HUNT, Andrew; JEFFRIES, Ron; KERN, Jon; MARICK, Brian; MARTIN, Robert C.; MELLOR, Steve; SCHWABER, Ken; SUTHERLAND, Jeff; THOMAS, Dave.

Manifesto for Agile Software Development. Disponível em: <<http://agilemanifesto.org/iso/en/>>. Acesso em: 14 de junho de 2016.