

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA

FRANCISCO ALBERNAZ MACHADO VALÉRIO
THIAGO TOKIORI HIGUCHI

**CLASSIFICAÇÃO COMPORTAMENTAL DE CLIENTES DE LOJAS
VIRTUAIS A PARTIR DE PADRÕES DE NAVEGAÇÃO**

CURITIBA

2014

FRANCISCO ALBERNAZ MACHADO VALÉRIO
THIAGO TOKIORI HIGUCHI

**CLASSIFICAÇÃO COMPORTAMENTAL DE CLIENTES DE LOJAS
VIRTUAIS A PARTIR DE PADRÕES DE NAVEGAÇÃO**

Trabalho apresentado à disciplina de Trabalho de Conclusão de Curso 2 do Curso Superior de Bacharelado em Sistemas de Informação do Departamento Acadêmico de Informática – DAINF - da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação..

Orientador: Alexandre Reis Graeml

CURITIBA

2014

RESUMO

HIGUCHI, Thiago T., VALÉRIO, Francisco A. M.. CLASSIFICAÇÃO COMPORTAMENTAL DE CLIENTES DE LOJAS VIRTUAIS A PARTIR DE PADRÕES DE NAVEGAÇÃO. 40 f. – Departamento Acadêmico de Informática, Universidade Tecnológica Federal do Paraná. Curitiba, 2014.

O comércio eletrônico ganhou grande importância com a popularização da Internet e seu baixo custo operacional. Entretanto, apenas uma pequena quantidade dos acessos aos *sites* de *e-commerce* é convertida em compras. Baseado nisso, existe um grande interesse em técnicas para aumentar essa “taxa de conversão”. Uma das maneiras usadas para esse fim é a customização em massa, que busca oferecer uma experiência personalizada aos clientes. Para poder oferecer essa experiência a cada cliente, existem diversas técnicas baseadas em mineração de dados. Este trabalho tem como objetivo classificar usuários anônimos entre diferentes perfis já descritos na literatura com base em seus padrões de navegação, utilizando mineração de dados na web.

Palavras-chave: comércio eletrônico, loja virtual, mineração de uso, mineração web, intenção da compra

LISTA DE FIGURAS

FIGURA 1	– Fases do ciclo de vida estratégico	7
FIGURA 2	– Gráfico com o faturamento do comércio eletrônico no Brasil	8
FIGURA 3	– Gráfico com o número de consumidores do comércio eletrônico no Brasil	9
FIGURA 4	– Caso de <i>compra direta</i>	19
FIGURA 5	– Primeiro caso de <i>deliberação</i>	19
FIGURA 6	– Segundo caso de <i>deliberação</i>	20
FIGURA 7	– Terceiro caso de <i>deliberação</i>	20
FIGURA 8	– Primeiro caso de <i>aquisição de conhecimento</i>	21
FIGURA 9	– Segundo caso de <i>aquisição de conhecimento</i>	21
FIGURA 10	– Primeiro caso de <i>hedônica</i>	22
FIGURA 11	– Segundo caso de <i>hedônica</i>	22
FIGURA 12	– Terceiro caso de <i>hedônica</i>	22
FIGURA 13	– Diagrama com todos os <i>caminhos</i> para as IVs	23
FIGURA 14	– Representação do padrão MVC	28
FIGURA 15	– Representação sistêmica do funcionamento do protótipo	28
FIGURA 16	– Diagrama de caso de uso	29
FIGURA 17	– Diagrama de Sequência	30
FIGURA 18	– Resultados dos Testes	35
FIGURA 19	– Distribuição dos Papéis	35

LISTA DE TABELAS

TABELA 1	– Recursos de Hardware para Desenvolvimento	25
TABELA 2	– Recursos de Software	26
TABELA 3	– Resultados dos Testes Abertos	34

SUMÁRIO

1	INTRODUÇÃO	4
1.1	OBJETIVO GERAL	6
1.2	OBJETIVOS ESPECÍFICOS	6
1.3	JUSTIFICATIVA	6
1.4	MOTIVAÇÃO	9
1.5	CONTEÚDO DO PROJETO	9
2	REFERENCIAL TEÓRICO	11
2.1	MINERAÇÃO DE DADOS	11
2.2	MINERAÇÃO WEB	12
2.3	MINERAÇÃO DE USO	13
2.4	CUSTOMIZAÇÃO EM MASSA	14
2.5	CUSTOMIZAÇÃO EM MASSA NA <i>WEB</i>	14
2.6	PERFIS DE USUÁRIOS	15
2.7	EXPRESSÕES REGULARES	16
3	METODOLOGIA	17
3.1	CLASSIFICAÇÃO DA INTENÇÃO DA VISITA	17
3.1.1	Caminhos	18
3.1.1.1	Compra direta	19
3.1.1.2	Deliberação	19
3.1.1.3	Aquisição de Conhecimento	20
3.1.1.4	Hedônica	21
4	RECURSOS DE HARDWARE E SOFTWARE	25
4.1	RECURSOS DE HARDWARE	25
4.2	RECURSOS DE SOFTWARE	25
5	ESPECIFICAÇÃO DO SISTEMA	27
5.1	PREMISSAS	27
5.2	PADRÃO DE PROJETO	27
5.3	ARQUITETURA DO SISTEMA	27
5.4	DIAGRAMA DE CASO DE USO	28
5.5	DIAGRAMA DE SEQUÊNCIA	29
6	VALIDAÇÃO DO SISTEMA	31
6.1	ESPECIFICAÇÃO DOS TESTES	31
6.1.1	Testes em Ambiente Controlado	31
6.1.2	Testes Abertos	32
6.2	RESULTADOS DOS TESTES	33
7	CONSIDERAÇÕES FINAIS	36
7.1	TRABALHOS FUTUROS	36
	REFERÊNCIAS	38

1 INTRODUÇÃO

Existe uma constante reformulação dos ambientes competitivos que vem sendo impulsionada por fatores como a globalização de mercados, crescimento da interpenetração de economias e aumento da interdependência de agentes econômicos (BECK; DEMIRGÜCKUNT, 2006). Neste cenário, a Tecnologia da Informação (TI) desempenha um papel cada vez maior dentro das organizações e, desde o final dos anos 70, a troca eletrônica de dados, ou *Electronic Data Interchange* (EDI), tem diminuído a burocracia dos papéis e, conseqüentemente, facilitado as operações dentro de empresas e entre elas (KALAKOTA; ROBINSON, 2003).

Segundo Rao *et al.* (2003), durante a década de 90 a TI já recebia grandes investimentos, principalmente para a automatização de processos internos, como, por exemplo, folhas de pagamento, finanças, ordens de compras, controle de entregas, recursos humanos e manufatura. Ainda durante esse período o grande crescimento da Internet popularizou um meio de alcance global, com custos relativamente baixos, para a operação e comunicação de organizações.

A partir deste novo paradigma, organizações foram capazes de reformular o jeito de interagir com clientes, parceiros e fornecedores (RAO *et al.*, 2003). Estas reformulações, incentivadas pelos avanços da TI, se mostraram tão vantajosas que as organizações que mais utilizavam a Internet para conduzir seus negócios apresentaram maiores receitas (USSBAJ, 2000).

Juntamente com estas vantagens estratégicas trazidas pela Internet e pela TI como um todo, a facilidade de efetuar transações entre duas ou mais partes propulsionou o crescimento do *e-commerce*, ou comércio eletrônico, conforme apresentam Grandon e Pearson (2004) e Kalakota e Robinson (2003).

De acordo com Kalakota e Robinson (2001), existem três forças que fomentam o crescimento do *e-commerce*.

Primeiramente as **forças econômicas**, que são facilmente observáveis, contribuindo

para o baixo custo operacional, de infraestrutura, de transação de valores e de suporte ao cliente.

As **forças de tecnologia e convergência digital** referem-se à convergência de conteúdo e transmissão oferecida pela TI. Nesse sentido, a tecnologia da informação fornece as ferramentas necessárias para levantamento, processamento, manipulação e distribuição de dados de forma eficiente. Esta convergência facilita a coordenação, integração e tomada de decisões.

Finalmente, as **forças de marketing e interação com clientes** são potencializadas, uma vez que estes novos canais de comunicação proporcionam um novo meio de interação com os consumidores, tanto no *marketing* (MCKENNA, 1999), quanto no suporte e no acompanhamento após a compra (SINGH, 2002).

Nos últimos anos, todas as vantagens do comércio eletrônico, e, principalmente, sua popularização entre consumidores, têm trazido um grande número de novos entrantes buscando uma fatia deste mercado (JAVALGI; RAMSEY, 2001; SCHMITZ; LATZER, 2002). Com este aumento de competição, as lojas virtuais têm adotado diversas estratégias em busca de mais visibilidade *online*, maior número de clientes, maior fidelidade dos clientes e, finalmente, maior número de vendas, conforme apresentado por Schmitz e Latzer (2002).

Uma das estratégias para aumentar as vendas em *e-commerce* é a customização em massa, ou *mass customization*, descrita por Pierrakos *et al.* (2003), Mobasher *et al.* (2000), Silveira *et al.* (2001). Segundo Silveira *et al.* (2001), esta estratégia pode ser brevemente apresentada como sendo a capacidade de entregar produtos, ou serviços, para uma grande quantidade de clientes, com características personalizadas individualmente. A personalização em massa também pode ser empregada em outras áreas, como por exemplo na montagem de produtos em escalas industriais (SILVEIRA *et al.*, 2001). Pierrakos *et al.* (2003) apresentam diversas aplicações da customização em massa focadas no ambiente virtual, como, por exemplo, na customização de conteúdo, *layout* e diferenciação entre produtos.

Existem, também, ferramentas que visam atacar o problema da conversão de acessos em vendas, dentre as quais pode-se destacar o *Google AdSense* e a sugestão de produtos existente em *sites* como *Amazon*. Porém, estas duas ferramentas utilizam outras abordagens para conseguir a conversão. No caso do *AdSense*, o *engine* analisa o conteúdo do *site*, geolocalização do usuário e outros fatores para gerar um anúncio relevante aos interesses do usuário (DAVIS, 2006). Já a *Amazon* utiliza dados de vendas anteriores e outras informações do cadastro do cliente, como idade, localização e gênero, para criar uma lista de sugestões de produtos (SONG, 2005).

Além desses dois exemplos de ferramentas, também existem exemplos brasileiros. O *SmartShop*, da *Btarget*, que faz parte do grupo Buscapé, promete um aumento na taxa de conversão da loja virtual por meio de uma personalização dos itens sugeridos, que passariam a ser produtos com maior probabilidade de compra. O site da *Btarget* (www.btarget.com.br) não apresenta muitos detalhes quanto ao funcionamento do sistema, a única informação é que são utilizadas técnicas de *big data* e inteligência artificial.

Este trabalho apresenta um protótipo capaz de contribuir para a customização em massa, identificando a Intenção da Visita (IV) de cada usuário a partir da extração de dados durante sua navegação pela loja virtual.

1.1 OBJETIVO GERAL

Este projeto tem como objetivo geral desenvolver um protótipo capaz de ser adicionado em lojas virtuais para identificar a Intenção da Visita de cada usuário a partir de dados extraídos durante sua navegação. Os dados utilizados são coletados a partir do momento em que o usuário entra no *website* e são descartados assim que o usuário sai.

1.2 OBJETIVOS ESPECÍFICOS

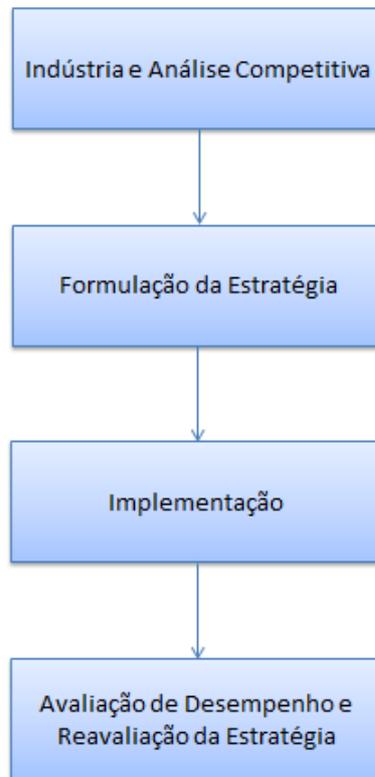
Foram definidos como objetivos específicos:

- Definir diferentes perfis de usuários a partir da literatura;
- Gerar Intenções da Visita a partir dos perfis identificados;
- Implementar um protótipo para identificar a IV do usuário de uma seção *web*;
- Avaliar a qualidade da classificação realizada a partir de testes do protótipo em um ambiente de vendas *online*.

1.3 JUSTIFICATIVA

Devido à crescente competitividade no *e-commerce*, diferentes estratégias vem sendo desenvolvidas não apenas com o objetivo de aumentar as vendas, como também pela simples necessidade de se manter no mercado (PLANT, 2000). De acordo com Turban *et al.* (2002), a teoria de mercado tradicional, tal qual a teoria de ciclo de vida estratégico, também se aplica

a lojas virtuais. A Figura 1 representa as fases para a criação de um ciclo de vida estratégico inicial.



Fonte: adaptado de Turban *et al.* (2012)

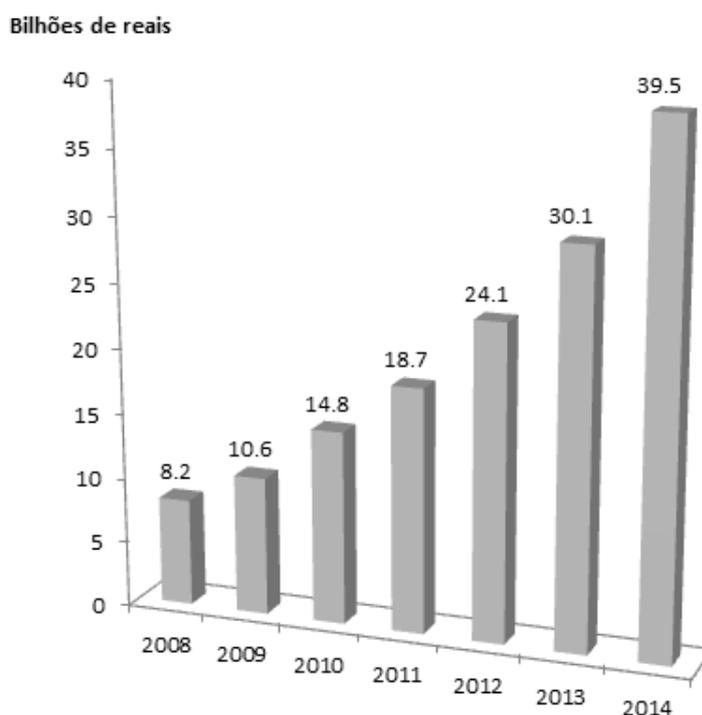
Figura 1: Fases do ciclo de vida estratégico

Na fase de formulação estratégica do ciclo de vida estratégico, diversas técnicas podem ser empregadas. Plant (2000) afirma que teorias tradicionais de *marketing* e de mercado podem ser aplicadas nesta etapa. Pelo fato de a Internet criar uma plataforma relativamente nova para o comércio, técnicas de *marketing* utilizadas em mídias tradicionais precisam evoluir para que tenham sucesso (HOFFMAN; NOVAK, 1997). Hoffman e Novak (1997) ainda apresentam abordagens mais detalhadas e apontam uma das mudanças fundamentais para que o *marketing* seja eficaz em ambientes virtuais, a individualização do consumidor.

Para uma experiência de “um-para-um” no ambiente virtual, métodos tradicionais, como, por exemplo, aplicação de questionários, se tornam um fracasso, uma vez que na grande maioria dos casos, devido a diversos fatores, o consumidor não se sujeita a respondê-los (HOFFMAN; NOVAK, 1997). Por esta razão, a extração de informações sobre o cliente enquanto ele navega se torna muito interessante para que experiências individualizadas e direcionadas ocorram em tempo real. O protótipo desenvolvido nesse projeto atuará

especificamente nesta função, extraindo dados e entregando informações que auxiliem o comércio eletrônico a atender os cliente de forma personalizada.

Existem algumas empresas dedicadas a implementar soluções com objetivos similares ao protótipo desenvolvido neste trabalho, como, por exemplo, a *Btarget*, apresentada na introdução. Isso que indica que existe uma demanda real por ferramentas que auxiliem na customização em massa. De acordo com a relatórios da empresa Ebit (2013), a previsão de faturamento do comércio eletrônico no Brasil é de aproximadamente 39,5 bilhões de reais. para o ano de 2014 A Figura 2 apresenta o crescimento do faturamento do comércio eletrônico entre os anos de 2008 e 2014. Se comparado o valor total de 2008 e a previsão para 2014, o crescimento é de aproximadamente 481% ao longo de seis anos.

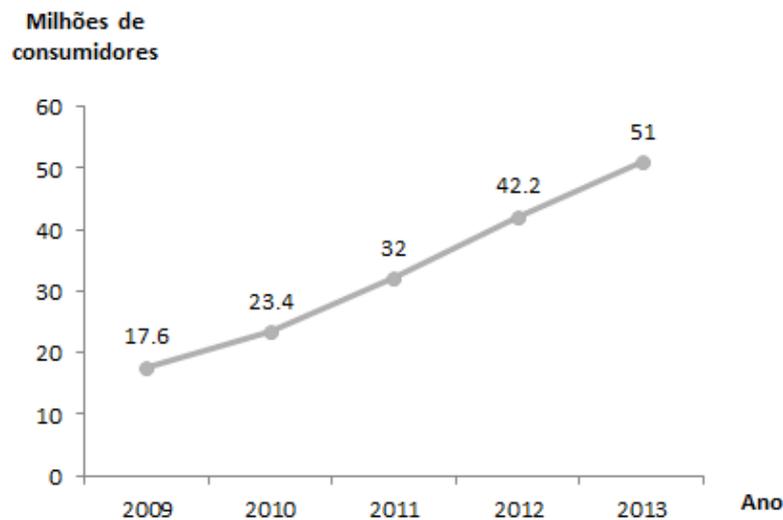


Fonte: adaptado de Ebit (2013)

Figura 2: Gráfico com o faturamento do comércio eletrônico no Brasil

Ainda segundo o relatório *Webshoppers* (EBIT, 2013), houve grande aumento no número de consumidores, que saltou de 32 milhões em 2011 para 42 milhões em 2012, sendo que a previsão para o ano de 2013 foi de 51 milhões. A Figura 3 apresenta os valores detalhados da pesquisa.

Os dados apresentados nas Figuras 2 e 3 apontam uma forte tendência de crescimento do comércio eletrônico no Brasil. A partir disso, iniciativas que melhorem os resultados de lojas virtuais se tornam ainda mais interessantes, tendo em vista que, além do número de



Fonte: adaptado de Ebit (2013)

Figura 3: Gráfico com o número de consumidores do comércio eletrônico no Brasil

consumidores estar aumentando, estes têm gastado mais em lojas virtuais.

1.4 MOTIVAÇÃO

Além da possibilidade de, a partir do protótipo desenvolvido, criar um produto com potencial para ser lançado no mercado, uma vez que seu nicho de atuação têm crescido consideravelmente nos últimos anos (EBIT, 2013), este trabalho permitiu integrar uma gama muito variada de conhecimentos adquiridos ao longo do curso, já que este serve como um dos requisitos para a obtenção do título de Bacharel de Sistemas de Informação, e também conhecimentos de outras áreas, como, por exemplo, *marketing* e psicologia.

1.5 CONTEÚDO DO PROJETO

A seguir é apresentada a organização do presente trabalho, destacando os principais assuntos abordados em cada capítulo:

- Capítulo 2.1 (Mineração de Dados): Esta seção apresenta detalhes quanto a mineração de dados, bem como seus diferentes tipos;
- Capítulo 2.2 (Mineração Web): Especificidades da mineração de dados no ambiente web são discutidos nesta seção;

- Capítulo 2.3 (Mineração de Uso na *Web*): As etapas para a mineração de uso na web são apresentadas nesta seção;
- Capítulo 2.4 (Customização em Massa): Detalhes da customização em massa no ambiente online são abordados nesta seção;
- Capítulo 2.6 (Intenção da visita dos usuários): As IVs identificadas na literatura são exploradas neste trecho do trabalho;
- Capítulo 3 (Metodologia): As etapas utilizadas para a execução do trabalho são descritas neste capítulo.
- Capítulo 4 (Recursos de Hardware e Software): Recursos necessários para o desenvolvimento do projeto.
- Capítulo 5 (Especificação do Sistema): Como o protótipo foi estruturado, nesta parte são apresentadas suas premissas e padrões.
- Capítulo 6 (Validação do Sistema): Como foram realizados os testes para a validação do protótipo e análise de seus resultados.
- Capítulo 7 (Considerações Finais): Por último, as considerações finais obtidas com o projeto e observações quanto a trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este capítulo contextualiza algumas áreas abordadas no projeto, como Mineração de Dados (MD), Mineração de Uso (MU), ou *Usage Mining*, Expressão Regular (ER), comércio eletrônico e customização em massa.

2.1 MINERAÇÃO DE DADOS

Transformar dados em conhecimento costumeiramente era feito por pessoas que analisavam os dados e davam sentido a eles, porém, em várias situações, a quantidade de dados é muito grande para ser analisada por pessoas. Além disso, a quantidade de dados a serem analisados cresce muito mais que a quantidade de pessoas para analisá-los. Por isso, Han e Kamber (2006) afirma que não vivemos na era da informação, mas na era dos dados. Assim surge a necessidade de criar uma maneira automática de extrair conhecimento de grandes volumes de dados (FAYYAD *et al.*, 1996b; TAN *et al.*, 2005; HAN; KAMBER, 2006).

Devido à grande coleta e acúmulo de dados que vêm ocorrendo em várias áreas, foram criadas novas teorias e ferramentas para extrair conhecimento desses conjuntos de dados. Assim surgiu o *Knowledge Discovery in Databases* (KDD – Descoberta de Conhecimento em Bancos de Dados) (FAYYAD *et al.*, 1996a).

O primeiro uso do termo KDD foi na *KDD Workshop* de 1989 (PIATETSKY-SHAPIRO, 1990). O KDD refere-se a buscar conhecimento em dados, de uma maneira geral. Por outro lado, a Mineração de Dados refere-se ao uso de algoritmos para buscar padrões nos dados, assim, a MD está inserida no KDD.

Uma das áreas que tem interesse na MD é o comércio eletrônico, que, além de gerar uma enorme quantidade de dados referentes às vendas e transações, também pode capturar a sequência realizada pelo usuário ao percorrer suas páginas, o chamado *clickstream*, ou sequência de cliques (KOHAVI *et al.*, 2004; KOHAVI, 2001). Além disso, no campo do comércio eletrônico, os dados são obtidos eletronicamente, sem necessidade de uma pessoa

coletá-los e inserí-los em um meio eletrônico para análise. Isso reduz consideravelmente o ruído existente nos dados, devido à eliminação do fator humano na digitalização dos dados (KOHAVI; PROVOST, 2001).

O processo do KDD envolve nove passos, conforme Fayyad *et al.* (1996a): primeiro, é necessário compreender o domínio de aplicação e definir os objetivos do KDD; então, deve-se criar o conjunto de dados que será analisado; em seguida, realizar a limpeza e o pré-processamento do conjunto; reduzir o número de variáveis analisadas; adequar os objetivos definidos durante a compreensão do domínio de aplicação a um método de MD específico, por exemplo: classificação ou agrupamento; escolher o algoritmo de MD e o método a serem usados; efetuar a mineração de dados; interpretar aquilo que foi minerado no passo anterior; e, finalmente, usar o conhecimento adquirido.

2.2 MINERAÇÃO WEB

A *web* tornou-se uma das maneiras mais utilizadas para buscar e compartilhar informação. Entretanto, nem sempre é fácil encontrar aquilo que se busca. Muitas vezes, existem muitos dados disponíveis, o que dificulta a busca por informações relevantes. Assim, a *web* tornou-se um ambiente interessante para a mineração de dados, no caso, chamada de Mineração *Web* (MW), que usa as técnicas de MD para encontrar e extrair informações relevantes de documentos e serviços existentes na *web* (KOSALA; BLOCKEEL, 2000).

Kosala e Blockeel (2000) estabelecem um paralelo entre o KDD e o MW, indicando que o último é o uma extensão do primeiro aplicado à *web*. Além disso, eles também dividem o processo de MW em quatro etapas, de maneira semelhante ao KDD: primeiro, é necessário buscar os recursos: recuperar os documentos que deseja-se analisar; em seguida, deve-se selecionar a informação e efetuar o pré-processamento de maneira automatizada a partir dos documentos recuperados no primeiro passo; então, deve-se generalizar, ou seja, descobrir padrões entre os documentos; finalmente, deve-se analisar os padrões encontrados.

Alguns trabalhos, como Cooley *et al.* (1997), dividem a MW em duas categorias: *Web Content Mining* (Mineração de Conteúdo *Web*), que se subdivide em abordagens baseadas em recuperação da informação e em bancos de dados, e *Web Usage Mining* (Mineração de Uso - MU).

A Mineração de Conteúdo *Web* busca informações relevantes no conteúdo da *web*, podendo buscar, além de HTML, documentos de outros formatos que estejam disponíveis, como PDF, doc, etc. O principal uso da abordagem de recuperação da informação é para encontrar ou

filtrar informações. Já a abordagem baseada em bancos de dados busca modelar e integrar os dados disponíveis para uso em futuras consultas (COOLEY *et al.*, 1997).

A Mineração de Uso *Web* tem como objetivo analisar as interações dos usuários com os *sites*, utilizando somente os dados gerados pelos navegadores durante a navegação (COOLEY *et al.*, 1997).

Outros trabalhos, como Kosala e Blockeel (2000), acrescentam uma terceira categoria: *Web Structure Mining* (Mineração de Estrutura *Web*). Essa modalidade de mineração busca descobrir a estrutura por debaixo dos enlaces existentes na *web*. Podendo ser usada para definir relações entre diferentes *sites*.

2.3 MINERAÇÃO DE USO

Segundo Pierrakos *et al.* (2003), o interesse pela análise do comportamento dos usuários no ambiente virtual tem crescido muito. Uma parcela da motivação para este episódio é a necessidade de entregar informação no momento correto para o usuário, que é constantemente bombardeado por uma grande quantidade de informação (*information overload*) (EDMUNDS; MORRIS, 2000). Assim, é importante mostrar ao usuário conteúdo que ele considere relevante, customizando as páginas pelas quais ele navega para mostrar esse conteúdo.

Pierrakos *et al.* (2003) propõem um procedimento para a mineração de uso na *web* com quatro fases, sendo elas:

- Coleta dos dados: o conteúdo e a estrutura de dados de diversas fontes (servidores *web*) são reunidos e armazenados em um banco de dados;
- Processamento dos dados: basicament, os dados já reunidos são consolidados, e possíveis inconsistências e ruídos são removidos;
- Identificação de padrões: informações são extraídas a partir dos dados. Diversas técnicas podem ser aplicadas, como, por exemplo, técnicas estatísticas ou identificação de padrão sequencial.
- Conhecimento de pós-processamento: com as informações já extraídas, ocorre a preparação para que estas sejam apresentadas de forma simples, como por exemplo a partir de relatórios e gráficos.

2.4 CUSTOMIZAÇÃO EM MASSA

A possibilidade de analisar o comportamento do usuário e, a partir disso, personalizar sua interação com o *website* é chamada *mass customization*, ou, simplesmente, customização em massa (MOBASHER *et al.*, 2000). Esta customização pode ser empregada em diversos casos. Muitos destes são amplamente discutidos por Pierrakos *et al.* (2003), como, por exemplo:

- **Memorização:** este refere-se a possibilidade de armazenar informações referentes ao usuário. Isto pode ser utilizado para analisar o comportamento ao longo do tempo. Esta funcionalidade normalmente é utilizada como parte de outra função para uma personalização mais completa;
- **Orientação (*guidance*):** no sentido de personalização, esta categoria é responsável por auxiliar o usuário a conseguir a informação que ele procura. Acredita-se que a orientação aumenta a lealdade enquanto diminui a sobrecarga de informação que costuma ser entregue para o usuário;
- **Customização:** a capacidade de personalizar conteúdo, estrutura e *layout* a partir dos dados de navegação do cliente. O objetivo da customização é basicamente melhorar a interação do usuário com o *website* como um todo;
- **Suporte de desempenho de atividade:** esta funcionalidade está relacionada com a execução de uma ação particular por parte do usuário. Sistemas deste tipo são classificados como sistemas adaptativos e comumente conhecidos como assistentes pessoais. De forma geral, o usuário é capaz de configurar alguns parâmetros, assim personalizando o modo de funcionamento de alguma ferramenta.

2.5 CUSTOMIZAÇÃO EM MASSA NA WEB

Devido à importância da personalização de serviços baseados na *web*, diversas técnicas foram desenvolvidas com esse fim (PIERRAKOS *et al.*, 2003), algumas delas são:

- (a) **Regras manuais para decisão:** os critérios para a personalização são criados a partir da intervenção de uma pessoa, muitas vezes chamada de especialista. Esta abordagem estática também pode ser obtida fazendo os usuários passarem por questionários que definem um dentre diferentes modelos de personalização.

- (b) Filtragem baseada no conteúdo: o objetivo desta técnica é aplicar métodos de aprendizagem de máquina para descobrir preferências do usuário. Exemplos desta abordagem utilizam padrões de navegação para construir e adaptar modelos de perfis usuários.
- (c) Filtragem colaborativa ou social: estas técnicas procuram personalizar um serviço apenas pesquisando por preferências em comum entre diferentes usuários que já estão armazenadas no sistema.

Diversos exemplos detalhados de cada uma destas abordagens são descritas por Lang (1995) e Manber *et al.* (2000).

2.6 PERFIS DE USUÁRIOS

Compras em lojas virtuais não seguem as mesmas regras das compras efetuadas em lojas físicas. Um exemplo disso é o custo envolvido para o cliente, muito menor no ambiente virtual. Para ir de uma loja a outra, basta clicar em outro *website*, não sendo necessário locomover-se fisicamente até outro local. Dessa maneira, o cliente tende a pesquisar mais antes de efetuar a compra (MOE; FADER, 2004). A partir disso, podemos concluir que somente aumentar a quantidade de acessos de um *website* não é o suficiente. Deve-se aumentar a taxa de conversão, isto é: a quantidade de acessos que resultam em vendas.

Devido a essas diferenças em relação às compras em lojas físicas, Moe e Fader (2001) analisaram os perfis de compra dos clientes de lojas virtuais. Concluíram que o processo de compra compreende as etapas de decidir entrar ou não no *site*; definir por quais páginas navegar; em seguida tem a decisão da compra; e, finalmente, comprar. Partindo dessas etapas, eles chegam a quatro perfis para os clientes de *e-commerce*, de acordo com o momento da compra (futura ou imediata) e o motivo da compra (planejada ou não):

- 1 - Visita para compra direta (*direct-purchase visit*): trata-se de um cliente que visita a loja virtual para uma compra já decidida. O cliente já está decidido a comprar o produto previamente escolhido. Portanto, é uma compra imediata e planejada;
- 2 - Visita para deliberação (*seach/deliberation visit*): o cliente visita a loja para reunir informações sobre o produto desejado e sobre o qual ainda está decidindo se vai efetuar a compra ou não. É o caso de uma compra futura e planejada;
- 3 - Visita para aquisição de conhecimento (*knowledge-building visit*): também é um caso no qual o cliente gosta do ato de visualizar os produtos. A diferença para a visita hedônica está no tempo da compra, neste caso, pode-se gerar uma compra futura e não planejada.

- 4 - Visita hedônica (*hedonic browsing visit*): é o caso em que o cliente gosta de visualizar os produtos, como um passatempo, mas não busca nada em específico. Nesse caso, podem ocorrer compras por impulso. Trata-se de uma compra imediata, mas não planejada;

Para este trabalho, cada um desses perfis de usuários foi mapeado para Intenções da Visita do usuário, mantendo a mesma nomenclatura. Isso deve-se ao fato de que o protótipo busca identificar a intenção do usuário ao visitar o *site*, e não categorizar o usuário em si.

2.7 EXPRESSÕES REGULARES

Expressões regulares permitem identificar cadeias de caracteres (*strings*) de maneira precisa e flexível. A expressão regular representa uma linguagem formal (conjunto de sequências de caracteres de um determinado alfabeto que é aceito por um autômato). Basicamente, a expressão regular especifica formalmente um padrão de strings (HOPCROFT *et al.*, 2003).

Uma expressão regular é composta por uma série de símbolos e caracteres especiais que formam uma sequência, também chamada de expressão. Um processador de expressões regulares deve interpretar a linguagem contida na expressão como uma regra, e irá compará-la à *string* de entrada (que será analisada), verificando esta cumpre todas as condições da expressão regular.

As expressões regulares são implementadas em linguagens de computação através de bibliotecas, como a RE2 (disponível em: <http://code.google.com/p/re2/>) para C++. Em outras linguagens, como Perl, as bibliotecas de expressões regulares fazem parte da biblioteca padrão da linguagem.

3 METODOLOGIA

Este capítulo apresenta uma visão geral das etapas realizadas durante o processo de desenvolvimento do projeto. Para atingir os objetivos do projeto foram identificadas as etapas descritas a seguir:

- Levantamento bibliográfico: estudo das áreas de conhecimento contidas no escopo do projeto;
- Pesquisa sobre ferramentas similares: identificação do estado da arte de ferramentas semelhantes ao protótipo proposto, bem como seus pontos fracos e fortes;
- Definição das possíveis intenções de visita: um apanhado aprofundado na literatura a respeito da intenção de visita de usuários em lojas virtuais;
- Planejamento do projeto: definição do escopo, criação do cronograma, detalhamento dos pacotes de trabalho, definição das intenções de visita que melhores se adaptam para o protótipo, levantamento de riscos, de recursos, de *hardware* e de requisitos e criação dos diagramas de classe, sequência e caso de uso do protótipo;
- Definição dos testes: estudo de quais testes podem ser empregados para a validação do protótipo;
- Implementação do protótipo: desenvolvimento do protótipo apresentado;
- Execução e análise dos testes: testes para a validação do protótipo.

3.1 CLASSIFICAÇÃO DA INTENÇÃO DA VISITA

Para classificar os perfis dos usuários dos *sites*, a equipe definiu que usaria *javascript* para coletar seus dados de navegação. Assim, foi definido que, para a primeira versão do sistema, seriam capturados os seguintes dados:

- (a) O usuário entra na página de um produto;
- (b) O usuário entra em uma página de categoria de produtos;
- (c) O usuário usa a ferramenta de pesquisa, buscando um produto específico. Neste caso, o usuário faz a busca usando o nome completo ou modelo do produto, de modo que a pesquisa terá poucos resultados (apenas um, na maioria dos casos);
- (d) O usuário usa a ferramenta de pesquisa, buscando um tipo ou categoria de produto. Neste caso, o usuário faz a busca usando um termo que não especifica um único produto. Por exemplo: “televisor”;
- (e) O usuário está na mesma página há muito tempo. O tempo para disparar essa condição é de um minuto. Esse tempo indica que o usuário pode estar fazendo uma leitura mais atenta da página;
- (f) O usuário entrou em uma página com produtos de uma categoria diferente da que estava visitando antes.

Assim, sempre que ocorre alguma dessas ações, um *cookie* é atualizado no lado do cliente. Nesse *cookie* existe uma *string* onde são guardados os dados de navegação do usuário usando um código para cada ação (no caso, os códigos vão de “A” até “F”). Assim, a cada nova ação, um novo código é concatenado à *string*.

Foi decidido usar uma *string* para armazenar as ações do usuário devido a sua simplicidade e facilidade de uso e, principalmente, porque *strings* são facilmente manipuladas usando expressões regulares.

3.1.1 CAMINHOS

Conforme mostrado na seção 2.6, existem publicações sobre perfis de compra de usuários em *sites* de vendas, entretanto, não foram encontradas descrições mais detalhadas sobre como o usuário de cada um desses perfis se comporta.

Assim, foram criados *caminhos*, que representam os passos que o usuário deveria realizar antes para cada um dos perfis identificados no capítulo 2 (*compra direta, deliberação, aquisição de conhecimento e hedônica*), que foram mapeados para as IVs.

Através de seções de *brainstorming* e da observação e conversas com usuários, foram definidos alguns *caminhos* para cada uma das IVs.

Nos diagramas que se seguem, um círculo pintado de preto indica início do *caminho*, dois círculos concêntricos, com o interno pintado de preto, indica final do *caminho*. Um retângulo com bordas arredondadas representa uma ação do cliente no *site*, e um retângulo sem bordas arredondadas indica a classificação da IV.

3.1.1.1 COMPRA DIRETA

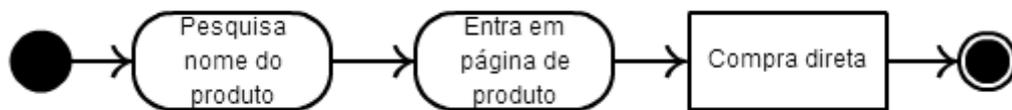


Figura 4: Caso de compra direta

Na Figura 4 é possível ver o *caminho* que leva à classificação do cliente como tendo Intenção da Visita de *compra direta*. Esse é o caso mais simples, pois o cliente já está decidido a realizar a compra: ele busca pelo nome do produto, entra em sua página e realiza a compra.

A expressão regular usada para mapear este *caminho* foi:

(CA\$)

Para um cliente com essa IV pode não ser recomendado realizar mudanças ou sugerir ofertas nas páginas nas quais ele navega.

3.1.1.2 DELIBERAÇÃO

Os casos de *deliberação* são mais complexos de se taxar, assim, foram necessárias várias regras para abranger todos os casos encontrados.

O primeiro caso é uma variação do *caminho* para a *compra direta* porém o cliente ainda fica algum tempo sem outras ações após entrar na página do produto, conforme pode ser visto na Figura 5. Isso indica que o cliente está com dúvidas quanto a adquirir o produto ou não.

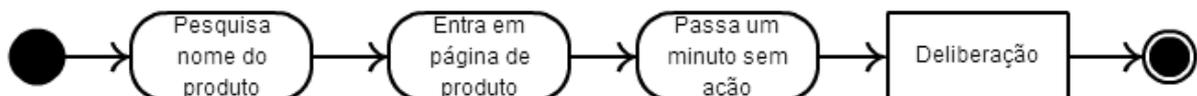


Figura 5: Primeiro caso de deliberação

O segundo caso é quando o cliente realiza uma pesquisa genérica (por alguma característica do produto, ou seu nome incompleto, por exemplo: “videogame”, ou “Nintendo”),

e entra na página do produto. Outra possibilidade é o usuário passar um tempo sem ação após entrar na página do produto. Isso pode ser visto na Figura 6.



Figura 6: Segundo caso de deliberação

O terceiro caso de *deliberação* é quando o cliente entra em uma página de categoria (por exemplo: página que lista os livros de Antropologia), passa, ou não, um minuto sem outra ação, e, finalmente, entra na página do produto. Este caso pode ser visto na Figura 7.



Figura 7: Terceiro caso de deliberação

Estes casos sugerem que o cliente busca o produto, mas ainda não está tão seguro quanto a comprá-lo, ou está em dúvida entre mais de um produto.

A expressão regular usada para mapear esta IV foi a seguinte:

$((DAE?\$) | (BAE?\$))$

3.1.1.3 AQUISIÇÃO DE CONHECIMENTO

Os casos de *aquisição de conhecimento* envolvem ciclos de mudanças de páginas de produtos e de categorias desses produtos.

A Figura 8 ilustra o caso em que o cliente realiza uma pesquisa genérica, então acessa a página desse produto e, a partir daí, muda de categoria até duas vezes e muda de página de produto até cinco vezes. Ou seja, o cliente pode pesquisar navegar por até cinco produtos da mesma categoria ou navegar por produtos de até duas categorias distintas.

A Figura 9 mostra com caso semelhante ao anterior, mas que inicia-se em uma página de categoria ou direto na página do produto, ao invés da busca genérica.

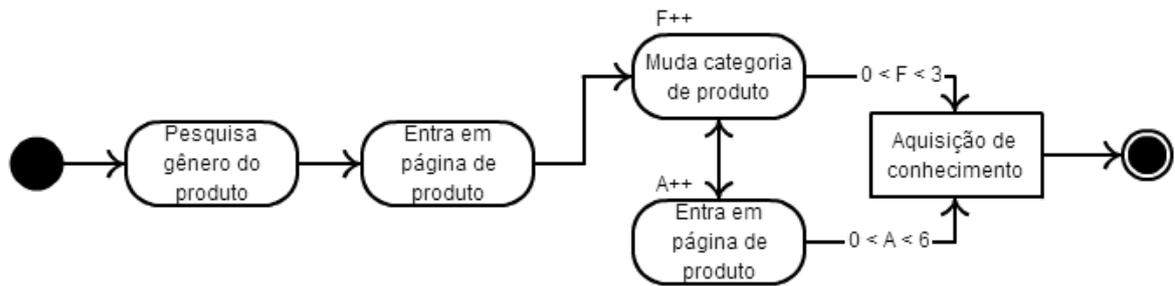


Figura 8: Primeiro caso de aquisição de conhecimento

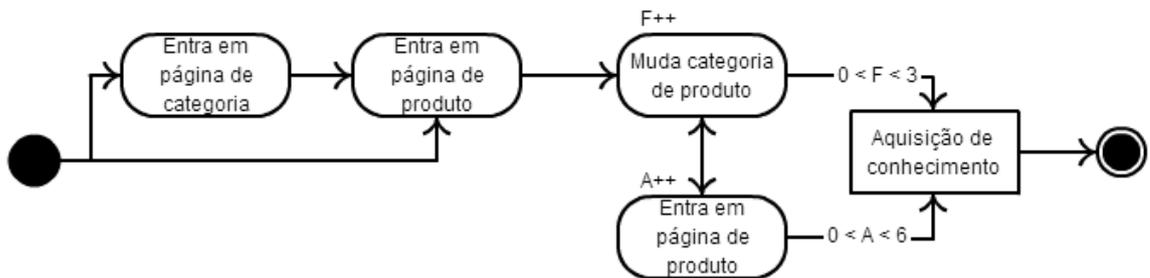


Figura 9: Segundo caso de aquisição de conhecimento

A expressão regular usada para mapear esta IV foi a seguinte:

$([BD]A*FA*FA*\$ | [BD]AF\{0,2\}AF\{0,2\}AF\{0,2\}AF\{0,2\}A\$ | [BD][AF]\$)$

Esta IV indica que o usuário está comparando dois ou mais produtos, provavelmente para adquirir mais informações sobre ambos.

3.1.1.4 HEDÔNICA

Finalmente, os casos de IV *hedônica* são o complemento dos casos de *aquisição de conhecimento*. Assim, no caso em que o cliente faz uma pesquisa genérica, entra na página do produto e, então, navega por mais de cinco produtos da mesma categoria ou por mais de dois produtos de categorias distintas, conforme a Figura 10.

Já a Figura 11 ilustra o caso em que o cliente entra direto na página do produto ou entra em uma página de categoria antes. Novamente, após esse primeiro passo, ele navega por mais de cinco produtos de mesma categoria ou por mais dois produtos de diferentes categorias.

O último caso de IV *hedônica* é quando o cliente entra duas ou mais vezes seguidas em páginas de categorias, conforme Figura 12.

A expressão regular usada para mapear esta IV foi a seguinte:

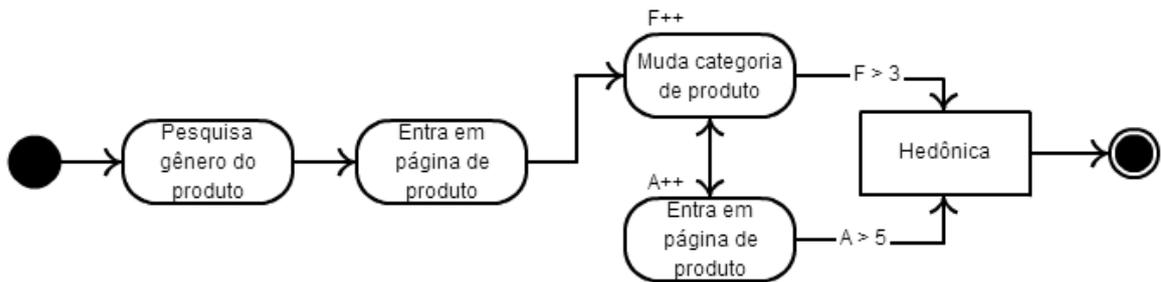


Figura 10: Primeiro caso de *hedônica*

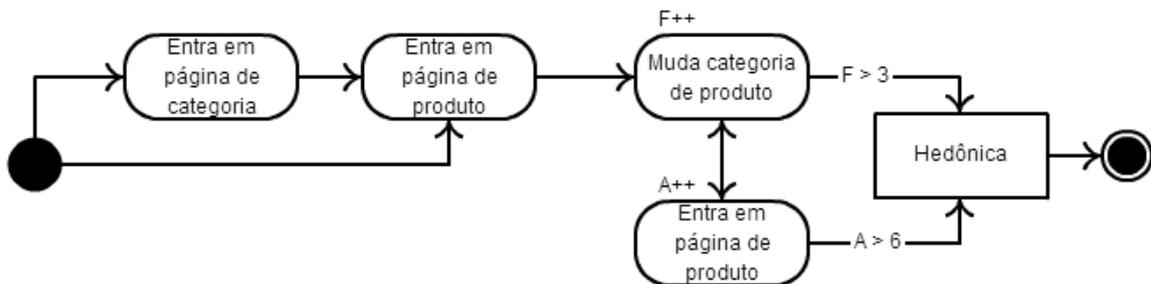


Figura 11: Segundo caso de *hedônica*

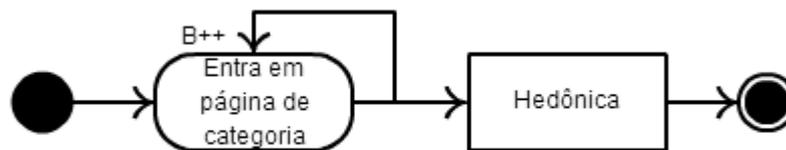


Figura 12: Terceiro caso de *hedônica*

$((B. \{0, 3\} B. \{0, 3\} B\$) | (FA*FA*FA*\$ | AF*AF*AF*AF*AF*AF*\$))$

Uma versão completa, incluindo todos esses caminhos visualizada na Figura 13.

Em versões futura deste projeto, o administrador do *e-commerce* que poderá escolher como será a heurística utilizada para definir quando o cliente será classificado em cada uma das Intenções das Visitas. Ainda assim, os *caminhos* usados neste trabalho podem ser melhorados a partir da observação dos padrões de comportamento dos clientes e podem ser apresentados como sendo a configuração padrão do sistema em sua versão comercial.

Conforme dito anteriormente, os caminhos possíveis para cada IV foram convertidos em expressões regulares. A *string* contida no *cookie* é passada como parâmetro para as funções de *regex* do PHP. Como já foi explicado, a função de *regex* busca na *string* de entrada sequências que se adequem aos padrões fornecidos, no caso, as expressões dos *caminhos* para as IVs.

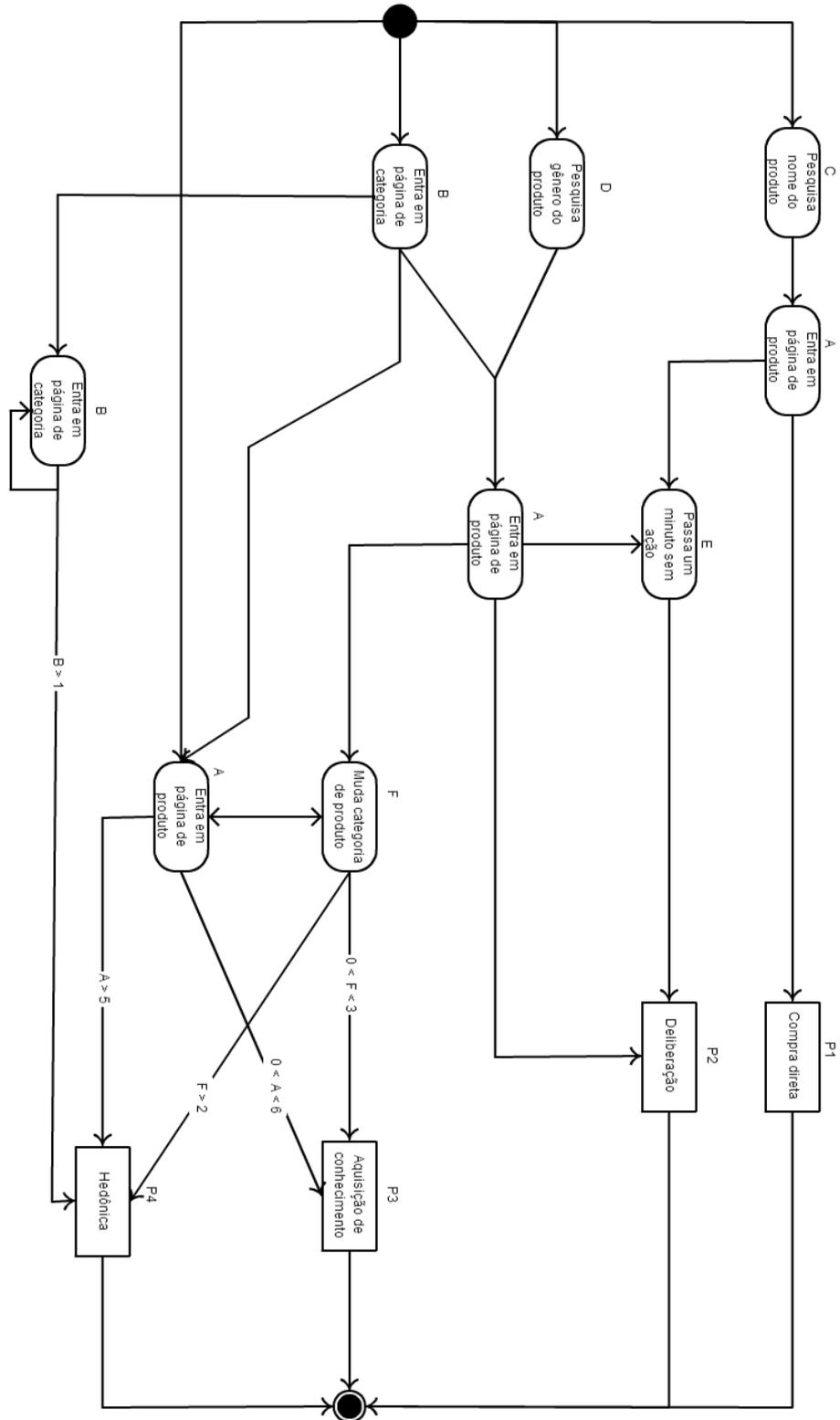


Figura 13: Diagrama com todos os caminhos para as IVs

Quando for identificada alguma sequência que se adeque a um padrão, a Intenção da Visita identificada para o usuário atual será salva em um *cookie*.

Note que a *string* pode conter mais de um padrão. Por isso, somente são considerados as sequências que ocorram no final da *string*. Dessa maneira, a Intenção da Visita definida para o usuário pode mudar caso ele mude seu comportamento.

4 RECURSOS DE HARDWARE E SOFTWARE

Este capítulo apresenta os recursos de hardware e software utilizados no projeto juntamente com suas respectivas aplicações.

4.1 RECURSOS DE HARDWARE

A parte de desenvolvimento e testes deste projeto contou, basicamente, com a utilização de dois computadores. A Tabela 1 apresenta detalhes da configuração de cada equipamento.

Tabela 1: Recursos de Hardware para Desenvolvimento

Hardware	Processador	Memória	Sistema Operacional
Computador 1	Intel Core i5-2400	8GB	Windows 7 Professional
Computador 2	Intel Core i5-2400	6GB	Windows 7 Ultimate

Em paralelo às máquinas utilizadas para o desenvolvimento, para os testes e a validação do protótipo foi cedido acesso a um servidor compartilhado que hospeda uma loja virtual real em funcionamento no Brasil. Por se tratar de um servidor compartilhado, detalhes em relação a hardware e software não são apresentados.

4.2 RECURSOS DE SOFTWARE

Os recursos de software levantados para o desenvolvimento deste projeto estão detalhados na Tabela 2. A definição destes recursos foi feita priorizando softwares de uso livre, desde que isso não influenciasse o desempenho do protótipo.

A linguagem de programação *javascript* foi escolhida devido a sua facilidade de ser embutida em *websites* de praticamente qualquer linguagem, como por exemplo: PHP, Ruby on Rails, Java Server Pages, HTML e HTML 5.

Tabela 2: Recursos de Software

Software	Versão	Licença	Utilização
Notepad++	6.5	Licença Pública Geral (GPL)	Ambiente de desenvolvimento
<i>Javascript</i>	1.8.5	Mozilla Foundation License	Linguagem de programação
Zoho	-	Free Edition	Cronograma e gerenciamento de atividades
PrestaShop	1.6	Open Software License	Testes e validação do protótipo

Durante a fase de validação do protótipo, a equipe conseguiu acesso a uma loja virtual do ramo de moda, em funcionamento desde 2013. Este *e-commerce* funciona a partir da plataforma de lojas virtuais *PrestaShop*. O *PrestaShop* é uma das soluções mais completas do mercado para a criação de lojas virtuais. Esta plataforma pode ser personalizada a partir de módulos e é distribuído sob uma licença *open source*. Mais detalhes referentes a esta plataforma podem ser encontrados em <http://www.prestashop.com/pt/>.

5 ESPECIFICAÇÃO DO SISTEMA

Este capítulo apresenta a especificação do protótipo desenvolvido neste trabalho.

5.1 PREMISSAS

Existem duas premissas básicas para o protótipo. A primeira é de que lojas virtuais sejam adaptadas para ter o *script* que recolherá os dados dos usuários e enviará para o servidor embutido no seu código fonte. A segunda premissa é de que as lojas devem ser responsáveis por receber os perfis do usuário e fazer mudanças em sua estrutura para melhor satisfazer os clientes.

5.2 PADRÃO DE PROJETO

Tendo como objetivo a melhor manutenção do código, o padrão de projeto *Model View Controller* (MVC) foi adotado. O MVC tem como característica a separação ente as camadas de “modelo”, “visão” e “controle” e a definição entre suas respectivas interações. De forma geral, a camada de visão é responsável pela apresentação dos dados para o usuário. A camada modelo engloba as estruturas de dados, incluindo classes e *structs*, por exemplo. E, finalmente, o controlador é responsável pela interação entra as outras camadas e validação das ações e operações (DEACON, 2009). A Figura 14 apresenta uma representação visual do MVC. Deacon (2009) apresenta informações mais detalhadas sobre este padrão.

Lindberg e Hallgren (2003) apresentam uma versão do MVC para uma arquitetura online, que melhor se adapta a esse projeto.

5.3 ARQUITETURA DO SISTEMA

A estrutura do protótipo desenvolvido neste projeto pode ser visualizada através de sua representação sistêmica na Figura 15. Basicamente, o protótipo é constituído de duas partes:

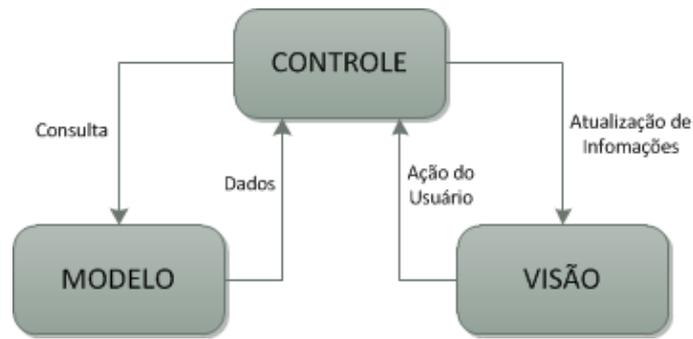


Figura 14: Representação do padrão MVC

a primeira é um *script*, que será embutido no *website* da loja virtual e dispara os *s* a partir do momento que o cliente executa determinada ação, e a segunda, responsável por receber informações sobre os clientes do *website*, processar estes dados, e, quando possível, informar a Intenção da Compra para o servidor do *e-commerce*.



Figura 15: Representação sistêmica do funcionamento do protótipo

Esta arquitetura utilizando um *script* no *website* da loja virtual e um “mini” servidor em PHP no próprio servidor para realizar o processamento dos dados foi definida por permitir que as regras de identificação da ICs estejam no mesmo local onde o *website* do *e-commerce* é mantido.

5.4 DIAGRAMA DE CASO DE USO

A partir das necessidades encontradas no projeto, o diagrama de caso de uso apresentado na Figura 16 foi criado. Os atores são o cliente, que, neste caso, representa o próprio consumidor que está acessando a loja virtual do *e-commerce*, e o *website*, que se trata do *web service* da loja virtual e o servidor que processará os dados coletados.

Existem cinco processos básicos no diagrama de caso de uso, sendo eles: “Entra em produto”, “Pesquisa específica” e “Pesquisa genérica”, “Entra em página de categoria” e “Muda de categoria”. Além destes, o processo “Passa tempo” foi adicionado ao diagrama, porém esta ação só é disparada internamente através do processo “Entra em produto”. Detalhes quanto a escolha de cada um destes “s” são apresentados na seção 3.1.

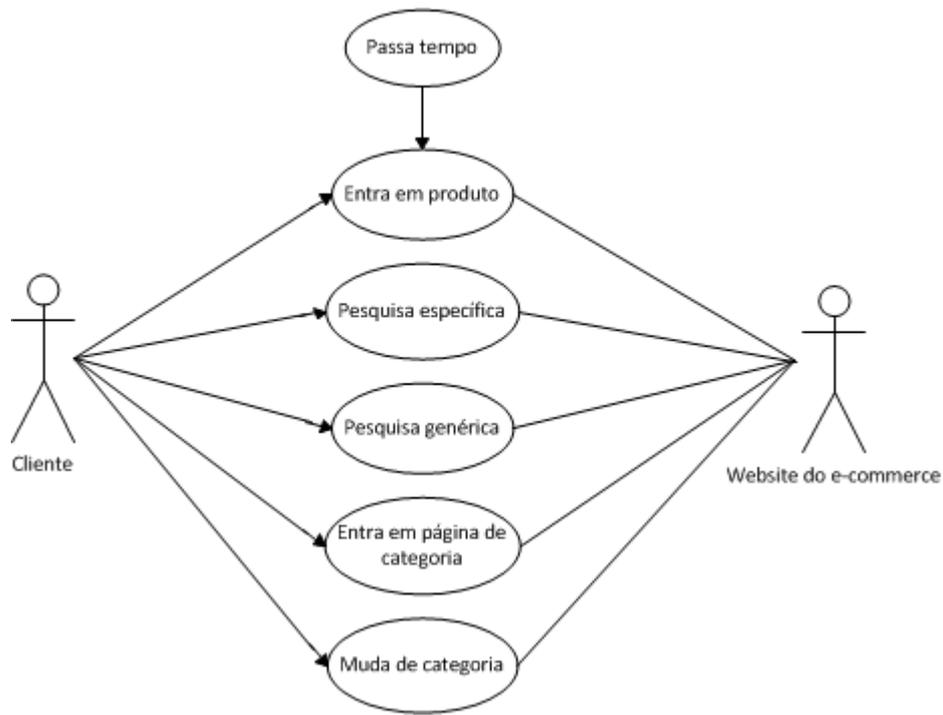


Figura 16: Diagrama de caso de uso

5.5 DIAGRAMA DE SEQUÊNCIA

A Figura 17 apresenta as interações durante o funcionamento do protótipo na forma de diagrama de sequência. O diagrama da Figura 17 representa o acionamento do gatilho de quando o cliente visualiza um novo produto, entretanto, o funcionamento dos outros gatilhos funcionam de forma semelhante, alterando-se apenas a página na qual o gatilho é acionado. A partir do acionamento de um dos gatilhos, uma requisição GET é gerada a partir do *javascript* no navegador do cliente, para o arquivo *servidor.php*, localizado no servidor do *e-commerce*, onde é feito o processamento e então, uma das IVs é identificada e salva nos *cookies*.

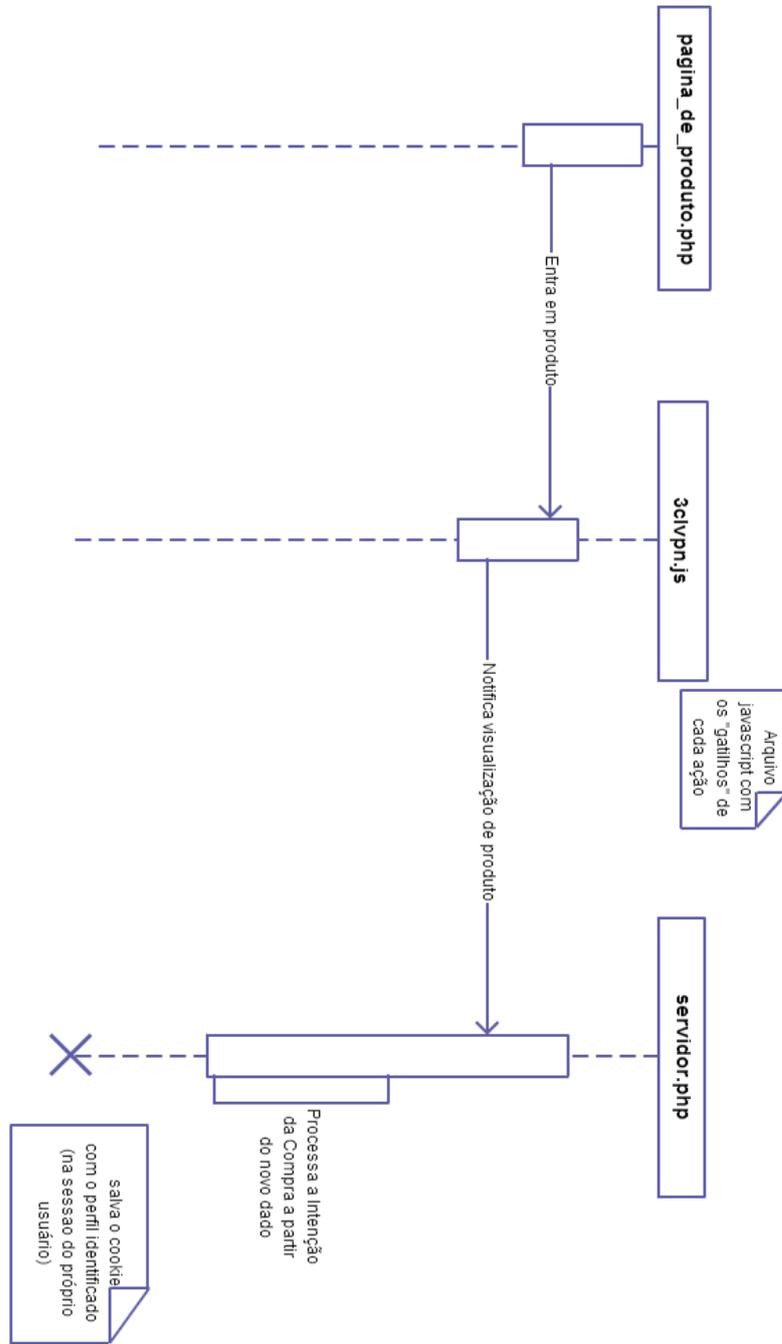


Figura 17: Diagrama de Sequência

6 VALIDAÇÃO DO SISTEMA

Neste capítulo estão especificados os testes realizados para verificar a eficiência do sistema proposto.

6.1 ESPECIFICAÇÃO DOS TESTES

Para a validação do sistema descrito neste trabalho, foram realizados dois tipos de testes distintos: um buscou verificar se o sistema identifica a intenção da visita do usuário correta em um ambiente controlado; e outro, em um *site* de uma loja, aberto a todos, que teve como objetivo verificar como se comportava o sistema em ambiente de produção.

6.1.1 TESTES EM AMBIENTE CONTROLADO

O teste em ambiente controlado consistiu em buscar voluntários e pedir a eles que interpretassem papéis relacionados ao uso de *sites* de compras. Alguns desses papéis são:

- Comprar um produto específico, com o nome e modelo;
- Navegar no *site* para passar o tempo;
- Pesquisar as diferenças entre dois produtos;
- Decidir a compra entre dois produtos;
- Pesquisar um produto para conhecê-lo melhor;

O roteiro básico para este tipo de teste foi o seguinte:

Os voluntários entram na sala com computadores com acesso ao *site* com o sistema instalado. Então, é informado a eles seus papéis. Neste momento, eles podem acessar o *site* da maneira como eles fariam normalmente. Não é dada a eles nenhuma outra instrução sobre como navegar no *site*. Quando os voluntários terminam seu acesso, eles são levados para fora

da sala. Após isso, a equipe analisa os resultados informados pelo sistema e verifica se eles conferem com o papel proposto aos voluntários.

A função principal deste teste foi verificar se os “caminhos de compras” que foram definidos para o sistema estavam de acordo com aquilo que é realizado por usuários reais. Assim, para isso, este teste foi realizado com 20 voluntários estudantes da UTFPR.

6.1.2 TESTES ABERTOS

O outro teste, em um *site* de uma loja virtual, foi realizado para verificar como o sistema se comporta em um ambiente de produção e se ele descobre a IV correta.

Para tanto, foi usado um *site* de compras já existente que permitiu a realização dos testes. Entretanto, não foi permitido realizar modificações no *site* principal, mas em uma cópia do mesmo. Assim, foi feita outra modificação no *site* para que, quando o usuário entrasse nele, surgisse um *pop-up* perguntando se ele desejava participar de um teste. Se o usuário aceitasse, ele era redirecionado a uma outra versão do *site*, já com o sistema instalado.

Ao redirecionar o usuário para o *site* de testes, outro *pop-up* informava qual deveria ser seu papel, de maneira similar ao que ocorria no teste anterior. Os papéis dos usuários eram escolhidos de maneira aleatória dentro de uma lista de oito possibilidades, todas relacionadas a uma das IVs tratadas pelo protótipo:

- 1 - Você deseja comprar a Camiseta Abercrombie & Fitch AF027.
- 2 - Você quer comprar uma camiseta Abercrombie tamanho G de cor escura.
- 3 - Você precisa de uma camiseta nova.
- 4 - Você deseja passar o tempo navegando pelo site.
- 1 - Você deseja comprar o moletom Hollister feminino HOMF064.
- 2 - Decida qual dos dois moletons você compraria? (HOMM011 ou HOMM001)
- 3 - Você precisa comprar um moletom para o aniversário de seu amigo.
- 4 - Você não tem nada pra fazer e resolveu dar uma olhada no site.

Ao ser redirecionado e ter seu papel informado, o usuário deveria navegar como faria normalmente, de acordo com seu papel. Ao terminar sua navegação, ele deveria clicar no

botão “Finalizar teste”, que enviaria os dados do teste para uma planilha armazenada no *Google Drive*. Os dados que foram guardados dos testes foram: a duração do teste, a *string* de ações do usuário e a IV definida para ele. Assim, pudemos verificar em quais casos o sistema identificou corretamente a intenção da visita.

6.2 RESULTADOS DOS TESTES

Os testes foram realizados conforme definidos na seção anterior durante os meses de julho e agosto de 2014.

Durante os testes em ambiente controlado, o servidor no qual o *site* estava hospedado não suportava uma grande quantidade de acessos simultâneos. Por isso, quando em uma aula na UTFPR foi pedido para que os alunos presentes acessassem o *site* para realizarem o teste, houve lentidão na resposta do *site*, chegando a não carregar a página, em alguns casos. Devido a este problema, não foram realizados outros testes com vários usuários ao mesmo tempo. Entretanto, isso não invalida o propósito deste trabalho, uma vez que não foi o protótipo que ocasionou esta lentidão no servidor.

Após isso, o *site* foi modificado para a realização do segundo tipo de teste, na própria loja virtual.

O teste ficou disponível durante três semanas no *site* entretanto, após uma semana do início dos testes, o *site* passou a apresentar muita lentidão, chegando a não carregar algumas páginas em vários casos. Foi verificado junto aos administradores do servidor de hospedagem, e eles informaram que estavam investigando as causas do problema, entretanto, até o final do período de testes não houve mais resposta deles e o problema persistiu.

Foi verificado que não era o protótipo que estava deixando o *site* lento, pois não havia *loops* no código, nem vazamento de memória. E quando o protótipo foi desinstalado do *site*, os problemas persistiram.

Essa situação impactou os testes em duas formas: a primeira, e mais aparente, é que diminuiu consideravelmente a quantidade de usuários que realizou os testes. Os usuários não queriam navegar em um *site* que demorava muito para carregar as páginas.

O segundo problema está relacionado à maneira como os dados eram enviados à planilha que armazenava os resultados. Para diminuir a carga do servidor, os dados que seriam gravados na planilha era enviados do próprio computador do usuário que estava acessando o *site*, através de um *link* para o *Google Forms* que era preenchido e enviado automaticamente.

Entretanto, devido ao grande atraso nas respostas do servidor durante a navegação, em muitos casos o *link* para o *Google Forms* não continha todos os dados vindos do servidor, chegando a faltar até duas ações do cliente e, conseqüentemente, duas respostas do servidor. Assim, alguns dos dados podem estar incompletos, não condizendo com a real resposta do protótipo.

Na Tabela 3, pode-se verificar: o Papel Sorteado para o usuário (de 1 a 4, em referência ao tipo de IV que ele deveria interpretar); a *String* de IVs, que foram sendo definidas pelo protótipo a cada ação do usuário, assim, o último caracter da cadeia é o que indica a IV final encontrada (também de 1 a 4, nos casos em que há um 0, indica que não foi encontrada uma correspondência. Nestes casos, é considerada a correspondência imediatamente anterior); e a *String* de Ações, que mostra quais foram as ações do usuário (de A a F, seguindo a notação utilizada na seção 3.1.1).

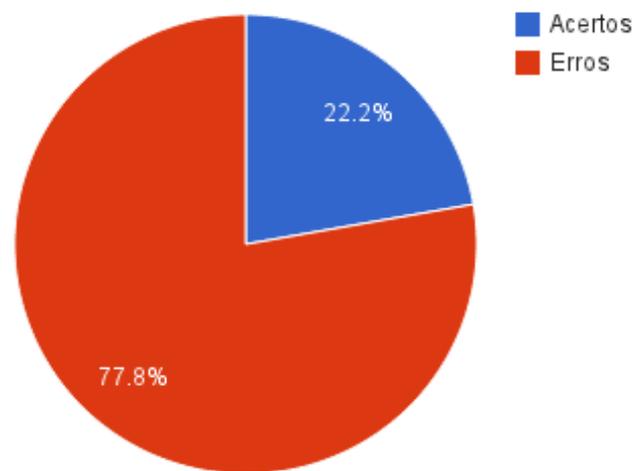
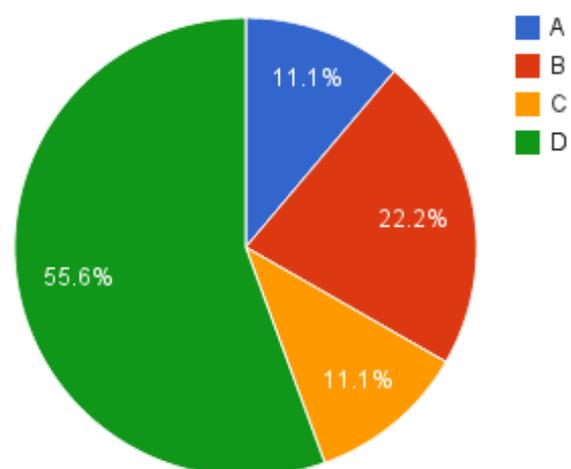
Das 19 respostas recebidas pelo *Forms*, somente 9 continham dados válidos. As outras 10 continham respostas repetidas ou faltavam dados, como, por exemplo, a *string* contendo as ações do usuário.

Destes nove testes válidos, em dois casos o protótipo identificou corretamente a IV do usuário, conforme pode ser visto na Figura 18

Tabela 3: Resultados dos Testes Abertos

Papel Sorteado	String de IVs	String de Ações
4	4020330	BBAAAafb
4	42000	BAAEFB
3	40442000	BBBBAEE
2	400	AFBB
1	420	BAC
4	40224	BBAEFB
4	40044	AFBBBB
4	40020	AAFBAFB
2	401001	CCAECA

Em relação aos tipos de papéis representados pelos usuário nesses nove testes, conforme pode-se verificara na Figura 19, houve cinco testes onde o papel possuía IV como sendo Hedônica (4), dois com IVs Deliberação (2), um com a IV Compra Direta (1) e outro de IV Aquisição de Conhecimento (3). Isso pode ser explicado pela forma como os papéis foram distribuídos: através de uma função aleatória.

Acertos dos testes**Figura 18: Resultados dos Testes****Tipos de papéis****Figura 19: Distribuição dos Papéis**

7 CONSIDERAÇÕES FINAIS

A partir das informações apresentadas na seção 6.2, pode-se observar diversos pontos-chaves em relação a este projeto como um todo.

Primeiramente, a taxa de acertos, ou seja, o número de vezes em que o protótipo chegou ao resultado esperado de acordo com o papel requisitado, obtida durante a fase de validação do sistema foi de 22.2% (discutido na seção 6.1.2). Entretanto, alguns fatores observados podem ter contribuído para esta taxa de acerto da identificação da IV abaixo do que era esperado.

Um destes fatores foi a alta latência observada no *website* que cedeu espaço para que o protótipo desenvolvido neste trabalho fosse testado. Esta latência dificultou consideravelmente a navegação no *website*. Esta era uma reclamação recorrente por parte de usuários que entravam em contato com o *website*. Acredita-se que este tenha sido um dos motivos pelo baixo número de “*input*” durante os testes.

Outra razão que pode ter contribuído para a taxa de acertos obtidas foi a elevada ocorrência de dados do onde a IV sorteada aleatoriamente era do tipo Hedônica. Estes papéis representaram mais de 55% de toda a amostra. Com isso, existe a possibilidade de que a heurística de caminhos utilizada para este caso não tenha a mesma eficácia que as outras, fazendo que a taxa de acerto obtida seja prejudicada.

Entretanto, apesar das dificuldades encontradas, a abordagem desenvolvida, utilizando expressões regulares para a identificação das Intenções das Visitas funcionou conforme o esperado. Além do mais, esta abordagem possui a característica de poder ser adaptada e ajustada para diferentes estruturas de comércios eletrônicos de forma simples, o que a torna muito interessante.

7.1 TRABALHOS FUTUROS

Como trabalhos futuros, fica claro que o protótipo desenvolvido possui diversos aspectos que ainda podem ser melhorados, como por exemplo:

- A implementação um tipo de gatilho genérico que possa ser embutido em qualquer página da loja virtual, fazendo com que o leque de passos analisados pelo site possa ser expandido sem dificuldades;
- A criação uma interface para que os administradores de lojas virtuais possam ajustar os *caminhos* de acordo com as necessidades de seu *e-commerce*, sem precisar editar arquivos fonte do protótipo;
- A partir do que já foi apresentado, juntamente de um estudo mais aprofundado a respeito do mercado, como por exemplo identificação de possíveis interessados e possíveis concorrentes, com as devidas modificações no protótipo para atender melhor o mercado, o artefato apresentado neste documento possui grandes chances de ser explorado comercialmente dentro de um nicho que está sempre disposto a absorver soluções que contribuam com seu crescimento.

REFERÊNCIAS

- BECK, T.; DEMIRGUC-KUNT, A. Small and medium-size enterprises: Access to finance as a growth constraint. **Journal of Banking & Finance**, v. 30, n. 11, p. 2931–2943, nov. 2006. ISSN 0378-4266. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0378426606000926>>.
- COOLEY, R.; MOBASHER, B.; SRIVASTAVA, J. Web mining: information and pattern discovery on the world wide web. In: **Ninth IEEE International Conference on Tools with Artificial Intelligence, 1997. Proceedings**. [S.l.: s.n.], 1997. p. 558–567.
- DAVIS, H. **Google Advertising Tools: Cashing in with AdSense, AdWords, and the Google APIs**. [S.l.]: "O'Reilly Media, Inc.", 2006. ISBN 9780596101084.
- DEACON, J. Model view controller (mvc) architecture. **<http://www.jdl.co.uk/briefings/MVC.pdf>**, 2009. Disponível em: <http://geographymyanmar.org/sites/default/files/gam/publication_file/MVC.pdf>.
- EBIT. Relatório web-shoppers. **São Paulo**, v. 28, 2013. Accessed: 2014-02-02. Disponível em: <<http://www.ebit.com.br/webshoppers>>.
- EDMUNDS, A.; MORRIS, A. The problem of information overload in business organisations: a review of the literature. **International Journal of Information Management**, v. 20, n. 1, p. 17–28, fev. 2000. ISSN 0268-4012. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0268401299000511>>.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, p. 37–54, 1996a.
- FAYYAD, U. M.; HAUSSLER, D.; STOLORZ, P. E. KDD for science data analysis: Issues and examples. In: **KDD**. [s.n.], 1996b. p. 50–56. Disponível em: <<http://www.aaai.org/Papers/KDD/1996/KDD96-009.pdf>>.
- GRANDON, E. E.; PEARSON, J. M. Electronic commerce adoption: an empirical study of small and medium US businesses. **Information & Management**, v. 42, n. 1, p. 197–216, dez. 2004. ISSN 0378-7206. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0378720604000205>>.
- HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. [S.l.]: Morgan Kaufmann, 2006. ISBN 9780080475585.
- HOFFMAN, D. L.; NOVAK, T. P. A new marketing paradigm for electronic commerce. **The Information Society**, v. 13, n. 1, p. 43–54, mar. 1997.
- HOPCROFT, J. E.; ULLMAN, J. D.; MOTWANI, R. **Introdução à Teoria de Autômatos, Linguagens e Computação**. Rio de Janeiro: Elsevier, 2003.

JAVALGI, R.; RAMSEY, R. Strategic issues of e-commerce as an alternative global distribution system. **International Marketing Review**, v. 18, n. 4, p. 376–391, ago. 2001. ISSN 0265-1335.

KALAKOTA, R.; ROBINSON, M. **E-business 2.0: Roadmap for Success**. [S.l.]: Addison-Wesley Professional, 2001.

KALAKOTA, R.; ROBINSON, M. **Electronic commerce**. John Wiley and Sons Ltd., 2003. Disponível em: <<http://dl.acm.org/citation.cfm?id=1074365>>.

KOHAVI, R. Mining e-commerce data: The good, the bad, and the ugly. In: **Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2001. (KDD '01), p. 8–13. ISBN 1-58113-391-X. Disponível em: <<http://doi.acm.org/10.1145/502512.502518>>.

KOHAVI, R.; MASON, L.; PAREKH, R.; ZHENG, Z. Lessons and challenges from mining retail e-commerce data. **Machine Learning**, v. 57, n. 1-2, p. 83–113, out. 2004. ISSN 0885-6125, 1573-0565. Disponível em: <<http://link.springer.com/article/10.1023/B%3AMACH.0000035473.11134.83>>.

KOHAVI, R.; PROVOST, F. Applications of data mining to electronic commerce. In: KOHAVI, R.; PROVOST, F. (Ed.). **Applications of Data Mining to Electronic Commerce**. Springer US, 2001. p. 5–10. ISBN 978-1-4613-5648-6, 978-1-4615-1627-9. Disponível em: <http://link.springer.com/chapter/10.1007/978-1-4615-1627-9_1>.

KOSALA, R.; BLOCKEEL, H. Web mining research: A survey. **SIGKDD Explor. Newsl.**, v. 2, n. 1, p. 1–15, jun. 2000. ISSN 1931-0145. Disponível em: <<http://doi.acm.org/10.1145/360402.360406>>.

LANG, K. NewsWeeder: learning to filter netnews. In: **in Proceedings of the 12th International Machine Learning Conference (ML95)**. [S.l.: s.n.], 1995.

LINDBERG, H.; HALLGREN, T. **Web-based architecture**. [s.n.], 2003. US Patent 20,030,028,540. Disponível em: <<http://www.freepatentsonline.com/y2003/0028540.html>>.

MANBER, U.; PATEL, A.; ROBISON, J. Experience with personalization of yahoo! **Commun. ACM**, v. 43, n. 8, p. 35–39, ago. 2000. ISSN 0001-0782. Disponível em: <<http://doi.acm.org/10.1145/345124.345136>>.

MCKENNA, R. Creating value in the network economy. In: TAPSCOTT, D. (Ed.). Boston, MA, USA: Harvard Business School Press, 1999. p. 145–158. ISBN 0-87584-911-3. Disponível em: <<http://dl.acm.org/citation.cfm?id=303444.303453>>.

MOBASHER, B.; COOLEY, R.; SRIVASTAVA, J. Automatic personalization based on web usage mining. **Communications of the ACM**, v. 43, n. 8, p. 142–151, 2000. Disponível em: <<http://dl.acm.org/citation.cfm?id=345169>>.

MOE, W. W.; FADER, P. S. Uncovering patterns in cybershopping. **California Management Review**, v. 43, n. 4, p. 106–117, 2001. Disponível em: <<http://www.rhsmith.umd.edu/faculty/wmoe/PDFs/MoeFader%20Cybershopping%20CMR%202001.pdf>>.

MOE, W. W.; FADER, P. S. Dynamic conversion behavior at e-commerce sites. **Management Science**, v. 50, n. 3, p. 326–335, 2004. Disponível em: <<http://mansci.journal.informs.org/content/50/3/326.short>>.

PIATETSKY-SHAPIRO, G. Knowledge discovery in real databases: A report on the IJCAI-89 workshop. **AI Magazine**, v. 11, n. 4, p. 68, dez. 1990. ISSN 0738-4602. Disponível em: <<http://www.aaai.org/ojs/index.php/aimagazine/article/view/873>>.

PIERRAKOS, D.; PALIOURAS, G.; PAPTAEODOROU, C.; SPYROPOULOS, C. D. Web usage mining as a tool for personalization: A survey. **User Modeling and User-Adapted Interaction**, v. 13, n. 4, p. 311–372, nov. 2003. ISSN 0924-1868, 1573-1391. Disponível em: <<http://link.springer.com/article/10.1023/A%3A1026238916441>>.

PLANT, R. T. **Ecommerce: Formulation of Strategy**. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000. ISBN 0130198447.

RAO, S. S.; METTS, G.; MONGE, C. A. M. Electronic commerce development in small and medium sized enterprises: A stage model and its implications. **Business Process Management Journal**, v. 9, n. 1, p. 11–32, 2003.

SCHMITZ, S. W.; LATZER, M. Competition in B2C e-commerce: Analytical issues and empirical evidence. **Electronic Markets**, v. 12, n. 3, p. 163–174, 2002. ISSN 1019-6781. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/101967802320245938>>.

SILVEIRA, G. D.; BORENSTEIN, D.; FOGLIATTO, F. S. Mass customization: Literature review and research directions. **International Journal of Production Economics**, v. 72, n. 1, p. 1–13, jun. 2001. ISSN 0925-5273. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925527300000797>>.

SINGH, M. E-services and their role in B2C e-commerce. **Managing Service Quality**, v. 12, n. 6, p. 434–446, dez. 2002. ISSN 0960-4529.

Zhengrong Song. **Methods and systems of assisting users in purchasing items**. mar. 2005. 6865546. U.S. Classification: 705/7.18; 705/14.35; 705/14.49; 705/26.7; 705/26.8 International Classification: G06F017/60. Disponível em: <<http://www.google.com.br/patents?id=3gUWAAAAEBAJ>>.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining, (First Edition)**. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005. ISBN 0321321367.

TURBAN, E.; LEE, J. K.; KING, D.; CHUNG; WARKENTIN. **Electronic Commerce: A Managerial Perspective**. 2nd. ed. [S.l.]: Prentice Hall Press, 2002. 2–20 p.

USSBAJ. Expansions in electronic commerce. **US Small Business Administration Journal**, jun. 2000.