



**ALEXANDER KIPPES**

**MINERAÇÃO DE DADOS E ANÁLISE DO  
COMPORTAMENTO DO CONSUMIDOR:  
ESTUDO DE CASO EM UM WEBSITE DE ENSINO  
A DISTÂNCIA**

**LAVRAS - MG  
2010**

**ALEXANDER KIPPES**

**MINERAÇÃO DE DADOS E ANÁLISE DO COMPORTAMENTO DO  
CONSUMIDOR: ESTUDO DE CASO EM UM WEBSITE DE ENSINO A  
DISTÂNCIA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências Programa de Pós-Graduação em Administração, área de concentração em Organizações, Mudanças e Gestão Estratégica, para a obtenção do título de Mestre.

Orientador

Dr. André Luiz Zambalde

Coorientador

Dr. Ahmed Ali Abdalla Esmín

**LAVRAS - MG  
2010**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca da UFLA**

Kippes, Alexander.

Mineração de dados e análise do comportamento do consumidor  
: estudo de caso em um website de ensino a distância / Alexander  
Kippes. – Lavras : UFLA, 2010.

116 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2010.

Orientador: André Luiz Zambalde.

Bibliografia.

1. Data mining. 2. Web mining. 3. Marketing. 4. Lato sensu a  
distância. 5. Ensino a distância. I. Universidade Federal de Lavras.  
II. Título.

CDD – 658.4033

**ALEXANDER KIPPES**

**MINERAÇÃO DE DADOS E ANÁLISE DO COMPORTAMENTO DO  
CONSUMIDOR: ESTUDO DE CASO EM UM WEBSITE DE ENSINO A  
DISTÂNCIA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências Programa de Pós-Graduação em Administração, área de concentração em Organizações, Mudanças e Gestão Estratégica, para a obtenção do título de Mestre.

APROVADA em 23 julho de 2010.

Dr. Gustavo Quiroga Souki	FUMEC
Dr. Luiz Marcelo Antonialli	UFLA
Dr. Ahmed Ali Abdalla Esmin	UFLA

Dr. André Luiz Zambalde  
Orientador

**LAVRAS - MG  
2010**

## **AGRADECIMENTOS**

À Universidade Federal de Lavras e ao Departamento de Administração e Economia, pela oportunidade concedida para a realização do mestrado.

Aos professores do Departamento de Administração e Economia da UFLA, pelos ensinamentos transmitidos e disponibilidade.

Ao professor Dr. André Luiz Zambalde, professor Dr. Ahmed Ali Abdalla Esmín e professor Dr. Luiz Marcelo Antonialli, pela orientação e apoio, fundamentais para a realização deste trabalho.

Ao meu amigo Jürgen, pelo apoio no último momento.

A Ana Cláudia, pela paciência, companheirismo, amizade, confiança e pela grande ajuda no trabalho.

Aos colegas, amigos e familiares que, de alguma forma, contribuíram para esta realização.

## RESUMO

O presente trabalho foi realizado com o principal objetivo de investigar o processo de mineração de uso web do curso de pós-graduação lato sensu a distância em Engenharia de Software Livre (ESL) da Universidade Federal de Lavras (UFLA), visando gerar informações e conhecimentos para a pesquisa, desenvolvimento e tomadas de decisão direcionadas a ações, práticas e atividades de marketing e relacionamento com o consumidor em organizações prestadoras de serviços educacionais na modalidade a distância. Quanto à metodologia, trata-se de uma pesquisa de natureza tecnológica, com abordagens qualitativa e quantitativa, com objetivos de caráter exploratório e descritivo, utilizando procedimentos de estudo de caso e de pesquisa operacional (análises matemáticas). Assim, além de entrevistas com o responsável pelo marketing da organização administradora do curso, acompanhadas de levantamento bibliográfico e documental, foram coletados e extraídos dados relacionados ao comportamento do navegador do visitante (usuário) em todas as páginas componentes da arquitetura do website “ESL”, utilizando ferramentas de software específicas. A abordagem de mineração de uso web permitiu a extração dos dados reais de navegação do usuário, ou seja, de todos os cliques, durante a entrada, a navegação e a saída do website. Na sequência, os dados passaram pelas fases intermediárias de limpeza, preparação e armazenamento, para posterior extração de padrões de navegação, análises matemáticas e geração de informações e conhecimento. Após a primeira etapa dos estudos, foi preliminarmente proposta esta reestruturação no website e, novamente, procedeu-se à estratégia de mineração de dados e descoberta de padrões de comportamento dos visitantes, com o objetivo de verificar o impacto da reestruturação realizada, em termos de retorno para a organização. Os resultados indicaram que o processo de mineração de dados do comportamento do usuário na web foi complexo e a organização não estava adequadamente preparada. Os desafios e os percalços encontrados chamaram a atenção para considerações gerais importantes, aplicadas a organizações que atuam no ensino a distância. O principal aspecto a se considerar é o adequado planejamento das relações entre marketing e relacionamento convencional, marketing e relacionamento web e equipe responsável pela tecnologia da informação. A busca da interação entre as equipes de marketing e relacionamento e os responsáveis pela tecnologia de web *mining* nas organizações é de fundamental importância para a elaboração de planos e atividades de propaganda e divulgação e, conseqüentemente, para a ampliação dos negócios. Conclui-se que os principais gargalos a serem tratados, considerando os contextos web *mining* de uso, comportamento do consumidor e marketing são: (1) as palavras chaves utilizadas na promoção dos cursos em ferramentas de busca; (2) o posicionamento sequencial das páginas do website,

visando criar um funil que leve o visitante a um formulário de inscrição ou de matrícula; (3) a arquitetura e o design adequados e atrativos do website; (4) a interação entre o marketing *offline* e o marketing *online* como, por exemplo, o Google Adwords e, finalmente, (5) a adequada escolha de páginas de referência ou de origem, tais como mídias sociais e outras páginas da web.

Palavras-chave: *Data Mining*. Web Mining usuário. O comportamento do consumidor. Marketing.

## ABSTRACT

The main and general objective of this research was investigating the Web Usage Mining process of the Latu senso on distance in Software Engineering Free (ESL), Federal University of Lavras - MG (UFLA), aimed at generating information and knowledge for research, development and decision-directed the actions, practices and marketing activities and customer relationships in organizations providing educational services in the distance mode. The methodology for this research is technological nature, qualitative and quantitative approaches, with goals of exploratory and descriptive, using a case study procedures and operations research (mathematical analysis). In addition the interviews with the responsible person of the marketing in the organization accompanied by bibliographic and documentary were collected and extracted data related to the visitor behavior on every page of the website architecture components ESL using specific software tools. There were collected and extracted related data to visitor browsing behavior (user) on every page of the website architecture components "ESL" in two specific time periods. The Web mining approach allowed the extraction of data from real user browsing, for any click during the entry, exit and navigation of the website. The data passed through the intermediate processes of cleaning, preparation and storage, for later extraction of browsing patterns, mathematical and statistical analysis and generation of information and so knowledge. After the first Period, was preliminarily proposed the restructuring of the website, and again proceeded a new strategy of data mining and visitors behavior discovery patterns in order to verify the impact of restructuring carried out in accordance return to the organization. The results indicated that the data mining process of the user behavior on the Web is complex and the organization was not properly prepared. The challenges and pitfalls encountered drew attention to important general considerations applied to organizations that work in distance education. The main aspect to consider is the proper planning of the relationship between conventional marketing, Web marketing and staff responsible for information technology. The search for interaction between teams and relationship marketing and those who are responsible for Web mining technology in organizations is crucial for the preparation of plans and activities and e-marketing, and consequently to the expansion of business. Anyway, the main bottlenecks to be addressed are considering the context of Web usage mining and marketing are: (1) the used key word in search engines, (2) the placement of sequential pages of the website aimed to create a funnel that takes the visitor on a registration form or registration, (3) the architecture and design appropriate and attractive website, (4) the interaction between offline marketing and online marketing, for example, Google AdWords, and finally, (5) the appropriate choice of reference

pages such as social media and other Web pages. Finally, it should be noted that practically all these bottlenecks are present in the context of the case study.

Keywords: Data Mining. Web user Mining. Consumer behavior. Marketing.

## LISTA DE FIGURAS

Figura 1	Descoberta do conhecimento em base de dados .....	26
Figura 2	Abordagens de <i>web mining</i> .....	27
Figura 3	<i>Web mining</i> de uso .....	28
Figura 4	Exemplo de registo de dados armazenados em cache web log .....	31
Figura 5	O funil de conversão .....	35
Figura 6	Fatores que influenciam o processo da decisão .....	39
Figura 7	Modelo processos de decisão do consumidor .....	51
Figura 8	O modelo do comportamento de um consumidor online .....	54
Figura 9	O processo de comunicação do comportamento do consumidor .....	57
Figura 10	Tela da ferramenta de análise da web Track4Mine .....	60
Figura 11	Sequência de comandos Java Script – Tags .....	61
Figura 12	Base de dados – coletados e tratados .....	62
Figura 13	Envio dos dados a uma base de armazenamento .....	62
Figura 14	Comportamento do usuário nas páginas do curso ESL .....	65
Figura 15	Tela do website Open Ufla .....	67
Figura 16	Tela do website do curso de Engenharia de Software (ESL) .....	68
Figura 17	Arquitetura em árvore do website ESL .....	70
Figura 18	Exemplo de acompanhamento de visitas no website ESL .....	72
Figura 19	Busca direta do Curso ESL na web – Google .....	73
Figura 20	Busca por palavras no curso ESL no Google – ineficiente .....	74
Figura 21	Páginas efetivamente monitoradas do website ESL .....	76
Figura 22	Uso da máquina de busca – Google .....	77
Figura 23	Gráfico das referências de entrada no website ESL .....	79
Figura 24	Agrupamento de cliques – Etapa I .....	82
Figura 25	Agrupamento por tempo – Etapa I .....	83
Figura 26	Sistemas operacionais utilizados para acesso a página ESL .....	84

Figura 27	Navegadores utilizados para acessar a página ESL .....	84
Figura 28	Resoluções mais utilizadas para acessar a página ESL.....	85
Figura 29	Estados de origem dos visitantes da página ESL.....	85
Figura 30	Arquitetura do website com formulário de matrícula e confirmação .....	89
Figura 31	Uso da máquina de busca – Google.....	90
Figura 32	Gráfico das referências de entrada no website ESL.....	93
Figura 33	Agrupamento de cliques – Etapa II.....	95
Figura 34	Agrupamento por tempo – Etapa II. ....	96
Figura 35	Sistemas operacionais utilizados para acessar a página ESL.....	97
Figura 36	Navegadores utilizados para acessar a página ESL .....	97
Figura 37	Resoluções mais utilizadas para acessar a página ESL.....	98
Figura 38	Estado de origem dos visitantes da página ESL. ....	99

## LISTA DE GRÁFICOS

Gráfico 1 Uso da máquina de busca – Google.....	77
Gráfico 2 Gráfico das referências de entrada no website ESL .....	79
Gráfico 3 Agrupamento de cliques – Etapa I.....	82
Gráfico 4 Agrupamento por tempo – Etapa I.....	83
Gráfico 5 Sistemas operacionais utilizados para acesso a página ESL.....	84
Gráfico 6 Navegadores utilizados para acessar a página ESL .....	84
Gráfico 7 Resoluções mais utilizadas para acessar a página ESL .....	85
Gráfico 8 Estados de origem dos visitantes da página ESL.....	85
Gráfico 9 Uso da máquina de busca – Google.....	90
Gráfico 10 Gráfico das referências de entrada no website ESL .....	93
Gráfico 11 Agrupamento de cliques – Etapa II .....	95
Gráfico 12 Agrupamento por tempo – Etapa II .....	96
Gráfico 13 Sistemas operacionais utilizados para acessar a página ESL.....	97
Gráfico 14 Navegadores utilizados para acessar a página ESL .....	97
Gráfico 15 Resoluções mais utilizadas para acessar a página ESL .....	98
Gráfico 16 Estado de origem dos visitantes da página ESL. ....	99

## **LISTA DE QUADROS**

Quadro 1 Descrição das páginas do ambiente do website ESL .....	70
Quadro 2 Palavras-chave Adwords usadas para acesso ao curso ESL. ....	78
Quadro 3 Palavras-chaves utilizadas na máquina de busca – modo orgânico. ...	78
Quadro 4 Exemplo da intercomunicação da população nesta pesquisa. ....	87
Quadro 5 Palavras-chave Adwords usadas para acesso ao curso ESL .....	91
Quadro 6 Palavras-chaves utilizadas na máquina de busca – modo orgânico. ...	92
Quadro 7 Exemplo da intercomunicação da população nesta pesquisa.....	101

## LISTA DE TABELAS

Tabela 1	Taxa de retorno - Etapa I .....	80
Tabela 2	Exame da atividade - Etapa I. ....	86
Tabela 3	Exame da atividade - Etapa I. ....	87
Tabela 4	Matriz de frequência na próxima fase - Etapa I. ....	88
Tabela 5	Taxa de retorno - Etapa II.....	94
Tabela 6	Exame da atividade - Etapa II.....	100
Tabela 7	Taxa de retorno na Etapa II.....	100
Tabela 8	Matriz de frequência na próxima fase na etapa II. ....	102
Tabela 9	Análises de funil de grupos - elementos de referência fase - etapa II. ....	103
Tabela 10	Análises de funil de grupos - tempo total por clique - Etapa II. ....	104
Tabela 11	Análises de funil de grupos - tempo total por clique - Etapa II. ....	105
Tabela 12	Análises de funil de grupos - número de cliques - Etapa II. ....	106

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	17
<b>1.1</b>	<b>Contextualização e motivação</b> .....	18
<b>1.2</b>	<b>Questões norteadoras e escopo da pesquisa</b> .....	19
<b>1.3</b>	<b>Objetivos e desafios</b> .....	22
<b>1.4</b>	<b>Estrutura do trabalho</b> .....	23
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	25
<b>2.1</b>	<b>Mineração de dados e análise de uso da web</b> .....	25
<b>2.1.1</b>	<b>Mineração de dados na web (<i>web mininig</i>)</b> .....	26
<b>2.1.1.1</b>	<b>Coleta de dados</b> .....	30
<b>2.1.1.2</b>	<b>Preparação e transformação de dados</b> .....	31
<b>2.1.2</b>	<b>Métricas analíticas (web analítica)</b> .....	31
<b>2.1.2.1</b>	<b>Frequência</b> .....	32
<b>2.1.2.2</b>	<b>A taxa de retorno (bounce rate)</b> .....	32
<b>2.1.2.3</b>	<b>Taxa de clique por anúncio (<i>click trough rate</i>)</b> .....	33
<b>2.1.2.4</b>	<b>Cliques por segundo</b> .....	33
<b>2.1.2.5</b>	<b>Movimento entre páginas (estados)</b> .....	34
<b>2.1.2.6</b>	<b>Funil de conversão</b> .....	34
<b>2.1.2.7</b>	<b>Página de visualização por dia</b> .....	35
<b>2.1.3</b>	<b>Trabalhos relacionados</b> .....	36
<b>2.2</b>	<b>Comportamento do consumidor</b> .....	37
<b>2.2.1</b>	<b>Modelo de decisão do consumidor</b> .....	38
<b>2.2.1.1</b>	<b>Influências sociais</b> .....	40
<b>2.2.1.2</b>	<b>Influências internas</b> .....	43
<b>2.2.1.3</b>	<b>Influências situacionais</b> .....	50
<b>2.2.1.4</b>	<b>Processos de decisão do consumidor</b> .....	50

2.2.1.5	Necessidade .....	51
2.2.1.6	Pesquisa - busca por informação.....	52
2.2.1.7	Avaliação das possibilidades.....	52
2.2.1.9	Consumo e após consumo .....	53
2.2.2	O comportamento do consumidor online .....	53
3	<b>METODOLOGIA</b> .....	58
3.1	Tipo de pesquisa .....	58
3.2	Coleta e análise dos dados.....	59
3.2.1	Coleta, limpeza e filtragem .....	60
3.2.2	Transformação dos dados.....	63
3.2.3	Mineração de dados.....	64
3.2.4	Geração de conhecimento (análise – uso aplicado aos negócios) .....	65
4	<b>RESULTADOS E DISCUSSÃO</b> .....	67
4.1	A arquitetura do site ESL.....	69
4.2	Divulgação e comportamento do consumidor ESL .....	72
4.3	Web mining ESL – Etapa I.....	75
4.3.1	Exame da campanha publicitária do Google Adwords.....	77
4.3.2	Exame da página de referência .....	79
4.3.3	Exame taxa de retorno (bounce rate) .....	80
4.3.4	Agrupamento das variáveis do usuário .....	81
4.3.5	Exame da atividade .....	86
4.3.6	Frequência da próxima fase .....	87
4.4	Webmining ESL – Etapa II.....	89
4.4.1	Exame da campanha publicitária do Google Adwords.....	90
4.4.2	Exame da página de referência .....	92
4.4.3	Exame taxa de retorno (bounce rate) .....	93

<b>4.4.4</b>	<b>Agrupamento das variáveis do usuário .....</b>	<b>95</b>
<b>4.4.5</b>	<b>Exame da atividade .....</b>	<b>99</b>
<b>4.4.6</b>	<b>Frequência da próxima fase .....</b>	<b>101</b>
<b>4.4.7</b>	<b>Análises de funil em cima dos grupos .....</b>	<b>103</b>
<b>4.5</b>	<b>Discussão e propostas de estratégias – marketing e web mining.....</b>	<b>106</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>110</b>
	<b>REFERÊNCIAS .....</b>	<b>113</b>

## 1 INTRODUÇÃO

Neste trabalho, entende-se que o comércio eletrônico, o comportamento das organizações que vendem produtos ou prestam serviços na rede mundial de computadores e o comportamento de clientes (usuários) podem ser estudados e entendidos com o apoio das tecnologias e sistemas de informação e de técnicas avançadas de descoberta e análise em bases de dados.

A vanguarda no contexto de algumas dessas técnicas, denominadas de mineração de dados (*data mining*), está possibilitando às mais diversas organizações dar um grande passo à frente dos concorrentes, viabilizando a conquista de uma parcela maior de mercado.

Em particular, a necessidade de conhecimento do comportamento dos usuários que acessam páginas da web vem pressionando drasticamente as organizações para a utilização de formas mais elaboradas de tecnologias. Também, o binômio comportamento dos usuários e uso da web vem sendo tratado como elemento de interesse e estudos no contexto da administração e dos sistemas de informação.

Nesse sentido, abordagem de mineração de dados de uso da web (*web usage mining*), definida como a coleta e a análise de dados de padrões comportamentais do usuário em uma página web, aplicada a um site de uma organização de prestação de serviços educacionais, foi o objeto de investigação e estudo do presente trabalho.

Especificamente, neste primeiro capítulo, busca-se contextualizar o tema e sua importância, o problema de pesquisa, os objetivos gerais e específicos e a estrutura e a dinâmica do trabalho.

## 1.1 Contextualização e motivação

Alguns dos principais desafios enfrentados pelas organizações nos últimos anos, em busca da sobrevivência no mercado, dizem respeito à organização da produção, redução de custos e gestão de pessoas. Além desses, as organizações atuais necessitam, principalmente, melhorar a qualidade do atendimento e conhecer melhor seu cliente, ou seja, tentar interpretar seus objetivos, expectativas e desejos.

Uma das tecnologias atualmente utilizadas pelas organizações para se buscar o conhecimento do cliente é a mineração de dados (*data mining*). Segundo Braga (2005, p. 11), “a mineração de dados provê um método automático para descobrir padrões em dados, sem a tendenciosidade e a limitação de uma análise baseada meramente na intuição humana”.

Particularmente a tecnologia internet, mais especificamente os serviços relacionados a websites (conjunto de páginas de uma organização) da World Wide Web (WWW ou apenas web), vem desafiando as organizações a utilizarem formas mais elaboradas de obtenção e manipulação de dados sobre os usuários.

Enfim, enquanto as empresas tradicionais se valem de estudos relacionados à observação do comportamento dos consumidores nos pontos de venda, de pesquisas qualitativas e quantitativas, e de aplicação de técnicas estatísticas e de *data mining* em seus repositórios de dados, as empresas de negócio eletrônico podem se valer de métodos e ferramentas de *web mining*.

Nesse sentido, a análise das informações armazenadas nos *logs* (registros) de computadores, denominados servidores web, permite conhecer o comportamento dos usuários e torna possível potencializar estudos, planos e ações de relacionamento e vendas. Essas informações passam, cada vez mais, a ter importante papel na definição das estratégias empresariais, possibilitando um

grande passo à frente dos concorrentes e viabilizando a conquista de uma parcela maior do mercado, seja ele de produtos ou serviços, como, por exemplo, a oferta de cursos via web, particularmente os que se relacionem a atividades como inscrição, matrícula e ensino-aprendizagem a distância.

Segundo Machado e Becker (2007), a web, com sua crescente disseminação de informações e serviços, alcançou diferentes contextos de aplicação, em que se destacam, além do comércio eletrônico, as bibliotecas digitais e a educação a distância. Conforme os autores, a acessibilidade e a facilidade de uso de ferramentas para manipular os recursos da Web têm tornado essa tecnologia uma escolha para a educação a distância, tanto em seu contexto externo, considerando a difusão da oferta e os procedimentos de inscrição e matrícula, em que, por meio de logs e registros, pode-se buscar conhecer melhor o comportamento do usuário, quanto interno, considerando os ambientes de ensino-aprendizagem e o processo de se disponibilizar informações online, suporte para discussões e componentes de avaliação dos diferentes comportamentos, ações e formas de execução de atividades propostas no ambiente de ensino virtual.

## **1.2 Questões norteadoras e escopo da pesquisa**

O presente trabalho partiu da premissa de que as técnicas de *web mining* de uso, ou seja, as técnicas de coleta e análise de dados navegacionais do usuário em uma página web, podem ser utilizadas para auxiliar a compreender o seu comportamento, considerando a descoberta de padrões ou regularidades que descrevam seu perfil.

Estas informações podem auxiliar no projeto ou na remodelagem de sites web, no desenvolvimento de agentes que auxiliem a navegação ou no

aprendizado do usuário e, mesmo, em uma abordagem analítica e crítica do conjunto ou plano de atividades de relacionamento web de uma organização.

Diante de tais possibilidades, seria de total interesse, para as organizações de maneira geral e para aquelas que atuam com a prestação de serviços relacionados ao ensino a distância em particular, entender cada vez melhor o comportamento do usuário, quando de uma visita à página web para busca de informações ou matrícula em determinado curso oferecido na modalidade a distância.

Estão disponíveis estudos e ferramentas que, por meio da análise do arquivo de *log* dos servidores web, fornecem informações estatísticas e de mineração sobre comportamentos dos usuários como, por exemplo, número de acessos por páginas, tempo de permanência, tráfego total e origem das requisições. No entanto, especificamente, não foram encontrados estudos que fazem referência a páginas web de organizações prestadoras de serviços relacionados à oferta de cursos e ensino a distância.

Assim, as questões norteadoras da presente pesquisa podem ser formuladas como sendo as seguintes:

- a) que informações sobre comportamentos definidos pelos dados originários de processos de acesso e navegação de usuários (potenciais alunos) a um conjunto de páginas de um website de oferta de um curso de pós-graduação a distância podem ser coletadas?
- b) como utilizar essas informações para potencializar estudos e investigações relacionadas a ações, práticas e ou atividades de marketing e relacionamento com os clientes?

As referidas questões têm como unidade caso ou escopo o *website* do curso de pós-graduação lato-sensu em Engenharia de Software Livre (ESL), oferecido pela Universidade Federal de Lavras (UFLA), em parceria com a Fundação de Apoio ao Ensino Pesquisa e Extensão (FAEPE).

Na busca por novos alunos, a UFLA adota ações e práticas de marketing em formato convencional (jornais, revistas, rádio, televisão, correios, pop-cards e catálogos) e via web (Google Adwords e website do curso).

No caso específico do website em estudo, quando um usuário se interessa pelo curso, ele é direcionado à página deste curso, a qual oferece informações básicas e de procedimento para matrícula. Esses usuários do website podem navegar nas mais diversas páginas, uma ou várias vezes, ou entrar e sair considerando uma única visita. Eles têm, portanto, comportamentos diversos. Por isso, é fundamental a descoberta de informações a partir dos dados de navegação, visando gerar conhecimento sobre o comportamento dos mesmos.

Os dados relacionados às opções do usuário (cliques e navegação) são automaticamente guardados em arquivos de log do servidor. Com esses dados, centenas e milhares deles, a organização pode realizar a mineração e o tratamento envolvendo, por exemplo, aqueles relacionados a quantas e quais páginas um usuário visitou, quanto tempo ele ficou em determinada página e se ele navegou da maneira como a equipe projetou o site. Essas informações auxiliam o entendimento do seu comportamento e a implementação de estratégias, ações e práticas direcionadas à melhoria do desempenho de processos de relacionamento com o cliente da organização.

Para interpretar essa grande quantidade de dados de uso, deve-se, então, se utilizar modernas tecnologias de descoberta do conhecimento em bases de dados (*Knowledge Discovery in Database*, ou KDD) fundamentadas em algoritmos de mineração de dados. Trata-se, portanto, da mineração de uso web (*web usage mining*), campo de pesquisa focado no desenvolvimento de técnicas

e ferramentas para estudo do comportamento dos usuários durante suas navegações em um website.

A compreensão das preferências de navegação de um usuário, aqui tratada, é uma importante etapa para o estudo de qualidade, desempenho e implementação de ações e práticas de negócios, principalmente em sites de comércio eletrônico. Por meio dos padrões de acessos dos usuários é também possível reestruturar, personalizar e adaptar interfaces e, mesmo, estratégias comerciais. Não faz parte do escopo deste trabalho, no entanto, fazer um comparativo de técnicas de web de uso e nem avaliar ou melhorar o desempenho dos algoritmos e programas computacionais.

### **1.3 Objetivos e desafios**

Este trabalho foi realizado com o principal objetivo de investigar o processo de mineração de uso web do curso de pós graduação lato sensu a distância em Engenharia de Software Livre (ESL) da Universidade Federal de Lavras (UFLA), visando gerar informações e conhecimentos para a pesquisa, desenvolvimento e tomadas de decisão direcionadas a ações, práticas e atividades de marketing e relacionamento com o consumidor em organizações prestadoras de serviços educacionais na modalidade a distância.

No contexto específico, buscaram-se:

- a) investigar o ambiente e a documentação de marketing, e marketing de relacionamento com o cliente de uma organização prestadora de serviços de ensino a distância;
- b) estudar técnicas, práticas e métricas de mineração de dados web (*web mining*) adequadas à análise de padrões comportamentais de navegação dos usuários do caso estudado;

- c) analisar e discutir os elementos de conhecimento resultantes de todo o processo, visando à sua adequada aplicação e difusão no contexto das organizações e cursos oferecidos a distância.

O desafio, então, é conseguir extrair informação e conhecimento a partir dos logs de um website, mais especificamente o comportamento do navegador dos usuários, ou seja, por meio de interpretações sobre os padrões de navegação encontrados, deve-se classificar e rotular os usuários. Enfim, buscar uma maneira subjetiva de analisar essas pessoas.

Considera-se comportamento, no contexto do presente trabalho, o conjunto das ações e reações que podem ser observadas num indivíduo, estando este no ambiente web e em dadas circunstâncias. Fazer análise do comportamento é determinar as características e dimensões da ocasião em que este ocorre, identificar as propriedades da ação e propor ou definir as mudanças necessárias ao ambiente em estudo. O evento comportamental, aqui, é o produto da navegação do sujeito no website especificado, o registro de log.

#### **1.4 Estrutura do trabalho**

O presente trabalho está dividido em cinco capítulos, considerando-se esta Introdução, além da listagem das referências bibliográficas, ao seu final.

No segundo capítulo tem-se o **Referencial Teórico**. Buscou-se fundamentar conceitos como mineração de dados web e web analítica (métricas), visando à aplicação em estudos sobre comportamento do consumidor.

No terceiro capítulo apresenta-se a **Metodologia** utilizada, definindo tipo de pesquisa e procedimentos metodológicos. Descreve-se como foi realizada a pesquisa, como o website foi efetivamente observado e como as

informações pertinentes foram extraídas. No contexto global de descoberta de conhecimento, buscou-se explicar como os dados foram coletados, até se chegar aos resultados, úteis, revelados.

No capítulo quatro, **Resultados e discussão**, são apresentados e discutidos os resultados do trabalho, particularmente as métricas e os gráficos correspondentes, incluindo considerações sobre tecnologias utilizadas pelos usuários, tempos de navegação, modos de navegação, enfim, o comportamento do usuário no website.

Em **Considerações finais**, são delineadas as oportunidades geradas, considerando uma adequada exploração de conceitos de *web mining*, projetos de website e estratégias de análise de comportamento e e-marketing. Além disso, identificou-se a contribuição do presente trabalho e foram propostas abordagens de trabalhos futuros.

Finalmente, têm-se as **Referências bibliográficas**, completando e informando sobre os elementos de investigação utilizados, considerando livros, periódicos, dissertações e teses e publicações disponíveis na internet.

## 2 REFERENCIAL TEÓRICO

Neste capítulo são apresentadas, em síntese, as bases referenciais relacionadas às áreas em estudo e investigação. Essas áreas são: mineração de dados, mineração web, análise de uso e métricas e aspectos teóricos sobre o comportamento do consumidor. Estas bases referenciais constituem instrumento explicativo para o adequado entendimento de todo o trabalho.

### 2.1 Mineração de dados e análise de uso da web

Nas últimas décadas, as organizações têm armazenado enorme quantidade de dados que são coletados de várias formas (ZAIANE, 1999). Essa quantidade de dados, por si só, não cria mais informações ou conhecimento. Foi necessário encontrar técnicas que ajudassem a produzir informações úteis a partir dos dados coletados (OBERLE, 2000).

Para isso, a mineração de dados (*data mining*), parte do processo de descoberta do conhecimento em base de dados (*knowledge discovery in database* ou KDD) tem sido amplamente utilizada pelas organizações (ZAIANE, 1999). Portanto, utiliza-se mineração de dados para extrair informação útil de uma base de dados tratada ou transformada (Figura 1), visando outra tomada de decisão na ação de marketing.



Figura 1 Descoberta do conhecimento em base de dados  
 Fonte: Adaptado de Zaiane (1999)

### 2.1.1 Mineração de dados na web (*web mining*)

No caso específico de dados relativos ao comportamento de um usuário web, tem-se que informações relacionadas à sua navegação podem, por exemplo, potencializar estratégias de relacionamento e ou ações de marketing (OBERLE, 2000). Assim, tem-se o que se denominou *web mining* (HEIDRICHS; LIM, 2003).

O *web mining* agrupa, em três diferentes abordagens, um conjunto de ferramentas que, além de proporcionar a descoberta da informação relevante, permite mapear e analisar o padrão de acesso e armazenamento de informações na web. Essas abordagens são a mineração de conteúdo (*web content mining*), a mineração de estrutura (*web structure mining*) e a mineração de uso (*web usage mining*) da web, conforme ilustrado na Figura 2.

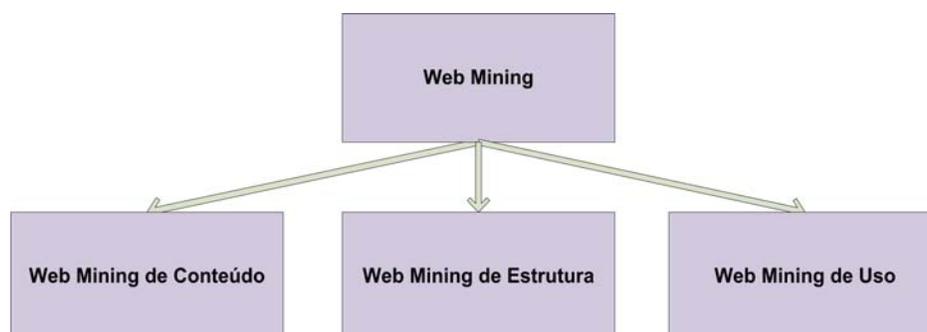


Figura 2 Abordagens de *web mining*  
Fonte: Rezende (2003)

De acordo com Pabarskaite e Raudys (2007), a mineração de conteúdo da web é a identificação da densidade, como, por exemplo, palavras especiais ou grupos de palavras em um website. A mineração de estrutura da web procura por domínios que apontam para um domínio específico. Essas técnicas podem ser úteis para explorar rankings dentro de uma máquina de busca ou as interconexões de um site na internet (REZENDE, 2003; ZAIANE, 1999).

A mineração de uso da web (*web usage mining* – *web mining* de uso) explora o comportamento dos visitantes, durante a navegação em páginas da web. A cada clique do usuário em uma página ou navegação em uma ou mais páginas, pode-se realizar o registro do caminho, o tempo e ou o local selecionado, em um arquivo de log. Esse registro de log pode ser também o número IP e o sistema operacional, entre outras informações técnicas úteis. Ele também pode oferecer informações sobre erros técnicos e de usabilidade de um website (PABARSKAITE; RAUDYS, 2007).

Tem-se, portanto, que a mineração de uso web pode ser utilizada para gerar dados relacionados ao “padrão de acesso geral” e à “personalização de uso” (Figura 3).

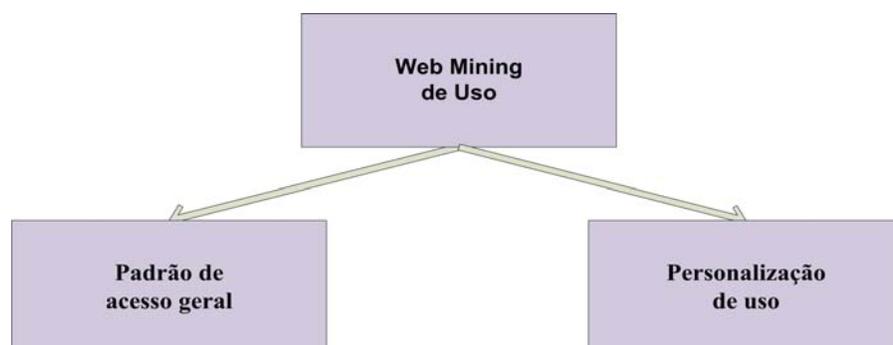


Figura 3 *Web mining* de uso  
Fonte: Rezende (2003)

Segundo Rezende (2003), servidores web armazenam dados em relação ao acesso a suas páginas de forma permanente. Os registros de acesso podem explicar o comportamento de usuários que buscam informações sobre determinado assunto e auxiliar na estruturação de um site, bem como em processos relacionados a marketing e a relacionamento com os clientes (usuários). “Efetivamente o interesse por este tipo de dado pode, por exemplo, determinar a vida útil de um produto (ou serviço), a estratégia associada a outro ou até mesmo alavancar um terceiro por meio de campanhas específicas” (REZENDE, 2003).

No caso de organizações de prestação de serviços educacionais via web (cursos), objeto de estudo deste trabalho, esse tipo de dado é fundamental para uma visão ampla e estratégica sobre os usuários do website, otimizando interfaces, atividades e negócios.

No âmbito da “determinação de padrões de acesso geral”, os registros são analisados de forma a apresentar uma tendência de acesso que pode levar à organização apropriada das páginas do website, visando eficiência e rapidez. Com o agrupamento de acessos representando transações web ou sessões de usuários (acesso às páginas do website durante uma visita), é possível obter um

conjunto de padrões de acesso valiosos para que as organizações entendam o comportamento de seus usuários. Algumas das técnicas de mineração de dados normalmente usadas para *web mining* são associação, sequência, classificação e agrupamento.

Com relação à “personalização de uso”, busca-se analisar tendências individuais de acesso. Nesse sentido, é possível a adequação dinâmica da estrutura do site, do formato dos recursos e da apresentação das informações ao longo do tempo, baseado no padrão de acesso de um usuário ao site.

A análise final dos padrões descobertos nos arquivos de registros (dados de coleta devidamente tratados) depende de uma série de estudos e ferramentas. Essas ferramentas visam adquirir conhecimentos que possam auxiliar na promoção da competitividade das organizações, principalmente considerando aspectos relacionados a estratégia de divulgação, elementos de relacionamento e navegação do usuário, entre outros aspectos organizacionais.

Conforme Holanda et al. (2006), em uma loja de varejo qualquer, o cliente caminha pelos corredores, observando produtos, comparando preços, selecionando determinados itens e descartando outros, até chegar ao caixa e registrar as suas compras. Nessa oportunidade, a referida loja pode estudar o seu comportamento e suas preferências. Já em um ambiente virtual, segundo o autor, “é possível rastrear o cliente mensurando o que ele observa, o tempo de observação e quais produtos são rejeitados depois de observados. Igualmente, ainda é possível reestruturar os corredores a fim de adequar aos anseios do cliente”.

Enfim, os dados comportamentais coletados podem alimentar uma base de dados de clique, gesto e trajetória, com informações tais como: Qual é o local do site mais visitado? E o menos visitado? Quais páginas do site são vistas por último, ou seja, onde os usuários encerram a sessão? Qual o perfil de navegação de um usuário cadastrado? E de um usuário rentável? Qual o perfil de

navegação de um usuário que cancela o serviço, faz reclamação ou processa a empresa? Quanto tempo os clientes gastam no site e quantas páginas eles visitam? Como induzir um usuário a se registrar, no intuito de obter informações valiosas sobre o mesmo?

Especificamente, do ponto de vista computacional, tem-se que, na mineração de uso web (*web mining* de uso), um servidor automaticamente coleta vários dados, tais como o IP do usuário (de onde ele veio?), data e hora (quando e quanto tempo ficou), tipo e versão do navegador (Internet Explorer, Mozilla, etc.), entre outros. Com esses dados, especialmente os *timestamps* (sequência de caracteres que denotam hora e data nos quais ocorreu determinado evento) e a identificação das páginas visitadas no site, tem-se que as análises de fluxo de cliques tornam possível descrever e registrar o tempo que um usuário gastou em cada página (IRENE et al., 2005; LEE et al., 2001).

Em síntese, tem-se que o processo técnico computacional de *web mining* de uso começa com a coleta, a preparação e a transformação dos dados, até se chegar à melhor técnica escolhida de mineração de dados e, finalmente, ao conhecimento. É esse conhecimento que pode garantir lucratividade às organizações que atuam no contexto do *e-commerce* e *e-learning*.

#### **2.1.1.1 Coleta de dados**

A coleta de dados (captura e armazenamento) é a base para os estudos relacionados com a web de uso, sendo importante escolher o método adequado. Segundo Kaushik (2007), quatro diferentes métodos são utilizados para captura ou coleta em sites web. O primeiro é a análise do arquivo de log de um servidor web; os demais são os métodos Java Script, Web Bacon e Sniffer. Desses, o método mais comum é o Java Script (KAUSHIK, 2007), que insere um conjunto

de instruções (um programa) em uma página que captura os dados de comportamento do usuário.

### 2.1.1.2 Preparação e transformação de dados

O arquivo de log gerado a partir da coleta tem um formato que não pode ser analisado diretamente (Figura 4).

```
198.102.031.003 - [10/Oct/2000:13:55:36 -  
0700] "GET /apache_pb.gif HTTP/1.0" 200 2326  
"http://www.ak-web4u.de/index.html "  
"Mozilla/4.08 [en] (Win98; I.;Nav) "
```

Figura 4 Exemplo de registro de dados armazenados em cache web log  
Fonte: Kaushik (2007)

São dados brutos que precisam ser limpos (processados) e transformados para que o software de mineração de dados possa interpretá-los. No caso específico da Figura 4, tem-se que o primeiro dado é o IP de número 198.102.031.003 (número do protocolo internet fornecido pelo provedor). Em alguns países, o IP indica a cidade, o estado e o país, e em outros países, somente o país de acesso para os usuários. O dado [10/Oct/2000:13:55:36-0700] indica data e hora em que o usuário navegou. O número 200 significa o estágio computacional da http-request e o número 2326 indica o número de bytes tratado. O dado <http://www.ak-web4u.de/index.html> é a página web de origem da entrada do usuário. Os demais dados são técnicos, como navegador e sistema operacional.

### 2.1.2 Métricas analíticas (web analítica)

As métricas analíticas são operações com dados de navegação, com

foco no acompanhamento do usuário no website. Pode-se calcular, por exemplo, casos de frequência de acesso em determinadas páginas, visitas versus tempo de acesso e as mais visitadas. Enfim, as métricas têm como objeto explicar ou permitir a análise de entradas, saídas e navegação no website. Na sequência, são apresentadas as métricas: frequência, taxa de retorno (*bounce rate*), *click trough rate* (taxa de atividade), cliques por segundo, número médio dos estados do visitante e páginas de visualização por dia.

### 2.1.2.1 Frequência

A frequência é a métrica que define o valor percentual de visitas por página versus o número de visitantes total. O visitante do site entra em uma página e com cada clique se faz uma ação da mudança de estado ou de página. As regras de associação são utilizadas para avaliar a porcentagem das pessoas que mudaram de um estado para outro estado. O pesquisador define os estados de uma arquitetura da web e conta a porcentagem de cada visita. Por exemplo, o visitante do site entra em uma página com três links, neste caso, cada link representa a possibilidade dele mudar de estágio, pois cada link representa um estado. Neste caso, a soma da frequência de estágios total tem de ser 100% (LEDFOORD; TYLER, 2007).

$$\text{FREQUÊNCIA(\%)} = \frac{\text{(NÚMERO DE VISITANTES POR ETAPA)}}{\text{(VISITANTES TOTAL)}}$$

### 2.1.2.2 A taxa de retorno (bounce rate)

A taxa de retorno é uma medida que expressa a porcentagem de visitantes que deixam a página após o primeiro clique, ou seja, a eficiência de

um website. Quando a taxa de retorno é grande, significa que a página de destino não é boa. Também, esta métrica dá uma visão geral sobre o cliente potencial e as pessoas que estão interessadas no website (LEDFOORD; TYLER, 2007).

$$\text{TAXA DE RETORNO (\%)} = \frac{\text{(NÚMERO TOTAL DE VISITAS QUE SAÍRAM)}}{\text{(ENTRADA TOTAL DA PÁGINA)}}$$

### 2.1.2.3 Taxa de clique por anúncio (*click trough rate*)

Em um website, os visitantes podem entrar e fazer um pedido a cada clique; eles podem visitar algumas páginas do site e outras, nem chegar a ver. Algumas dessas páginas têm a opção de anúncios. Em todas as páginas visitadas, o visitante pode realizar várias seleções ou opções por algum anúncio, realizando, por exemplo, uma opção de compra comercial. Ter um valor de volume de cliques sobre anúncios (*click trough rate*) é bom para posicionar um anúncio em determinado website ou determinada página. Um valor baixo de taxa clique por anúncio mostra que o *designer* da página e a página têm informações de pouco interesse ou os visitantes não estão vendo determinado anúncio (LEDFOORD; TYLER, 2007).

$$\text{CLIQUES POR ANÚNCIO (\%)} = \frac{\text{(CLIQUES)}}{\text{(ANÚNCIO)}}$$

### 2.1.2.4 Cliques por segundo

Conforme Menascé e Almeida (2000), clique por segundo é uma métrica, ou seja, o número de pedidos ao servidor, dividido por todo o tempo de

visita. O resultado mostra o tempo de uso do website para cada ação ou por página. O resultado pode significar para o analista que o visitante entrou e ficou lendo determinado assunto em uma página ou só viu os elementos não escritos e deixou o site ou página. No entanto, é necessário cautela porque esta métrica atende a outros elementos interativos, como anúncios.

$$\text{CLIQUES POR SEGUNDO} = \frac{(\text{NÚMERO SOLICITAÇÕES AO SERVIDOR})}{(\text{TEMPO TOTAL DE VISITA - SEGUNDOS})}$$

#### 2.1.2.5 Movimento entre páginas (estados)

Segundo Menascé e Almeida (2000), cada página da web ou grupo da website é um estado para o analista. Quando um usuário visita uma web site e se move entre ele, automaticamente se move entre os estados (entre páginas). Esta métrica dá ao analista uma visão geral do que os visitantes estão mais interessados.

$$\text{MOVIMENTO MÉDIO POR PÁGINA} = \frac{(\text{VISITANTE POR PÁGINA})}{(\text{NÚMERO TOTAL CLIQUES})}$$

#### 2.1.2.6 Funil de conversão

Funil de conversão (Figura 5) é uma série de páginas pelas quais deseja-se que um visitante passe antes de acessar a meta desejada pela organização. A finalidade de acompanhar essas páginas é ver a eficiência com que elas direcionam os visitantes para a meta definida. Se uma das páginas do funil for

muito complexa ou não tiver um projeto amigável, tem-se um número significativo de desistências e baixas taxa de conversão (alcance da meta). Podem-se acompanhar as taxas de desistências nas páginas que conduzem a uma meta utilizando, por exemplo, um relatório "visualização do funil". Conforme Eley e Tilley (2009), o objetivo em um site de *e-commerce* é, muitas vezes, a compra de um produto ou, no caso do presente estudo, a matrícula em um determinado curso.

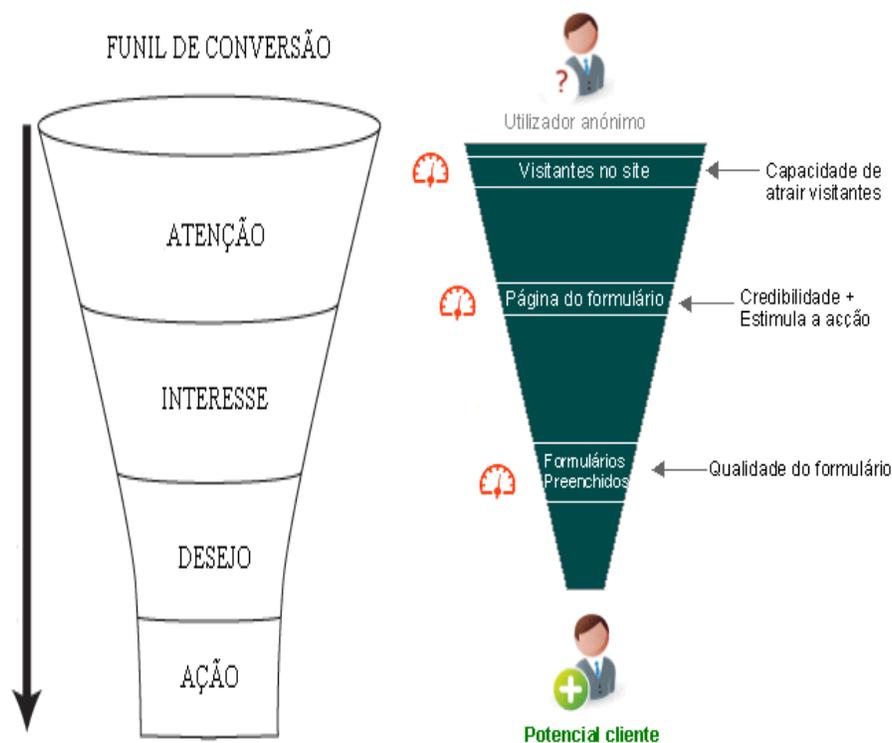


Figura 5 O funil de conversão  
Fonte: Eley e Tilley (2009)

### 2.1.2.7 Página de visualização por dia

Segundo Menascé e Almeida (2000), página de visualização por

dia, refere-se às visitas no dia e no número de páginas individuais servidas por dia. Quando uma empresa quer ver o sucesso de um determinado assunto ou tema que é colocado em uma página, este pode estar interessado no número de vezes que seu anúncio está sendo visto. Páginas populares podem exibir algumas centenas de milhões de páginas visitas por dia.

### **2.1.3 Trabalhos relacionados**

Na sequência, a título descrição de aplicações e estudos, são apresentados alguns trabalhos que relacionam *web mining* e técnicas web, visando à melhoria do relacionamento com o consumidor e atitudes de marketing. Entende-se que estes trabalhos relacionados são importantes para dar ao leitor uma visão de estudos e aplicações atuais e efetivamente direcionadas ao tema tratado nesta pesquisa.

Pabarskaite e Raudys (2007) descrevem como diferentes pesquisadores usam *web usage mining*. Eles explicam como funciona o processo de recolha e preparação da mineração de dados na web, como são usados em técnicas de mineração de uso da web de uma forma geral, como para identificar as sessões e evitar dados incorretos que podem levar a conclusões erradas. Técnicas básicas de *web usage mining*, como clustering, classificação e método de árvore na web são apresentados nesta publicação.

Ngai, Xiu e Chau (2009) descrevem uma abordagem especial para a gestão de relacionamento com clientes e as técnicas de mineração de dados que podem ser utilizados para criar, controlar e melhorar o relacionamento com o cliente. Este artigo fornece uma breve visão geral e compreensível de diversas técnicas de mineração de dados e os seus usos.

Tug, Sakiroglu e Aslan (2003) descrevem como analisar o comportamento do consumidor on-line por meio de arquivos de log de web.

Vários conceitos de mineração de dados são apresentados que podem ajudar a encontrar o conhecimento escondido em arquivos de log e assim contribuir para melhorar as decisões de marketing. Nesta pesquisa utiliza-se um algoritmo genérico que explora em que partes de um site os usuários são concentrados, resultando em um grupo de páginas.

Finalmente, Lee et al. (2001) apresentam um fluxo de visualização de clique que ajuda a interpretar o comportamento do cliente. Ele explica como interpretar o clique em blocos e que conclusões podem ser tiradas em relação ao comportamento do consumidor. Este artigo é um bom ponto de partida para a compreensão dos princípios básicos de interpretação e agrupamento de comportamento de diferentes usuários.

## **2.2 Comportamento do consumidor**

Segundo Engel, Blackwell e Miniard (2000), os estudos sobre o comportamento do consumidor tiveram início no fim da Grande Depressão, no Estados Unidos. Com as tecnologias de comunicação e transporte, potencializaram-se as relações comerciais nacionais e a necessidade de comunicação com os mercados internacionais. Nas décadas seguintes, a ciência buscou cada vez mais estudar o comportamento do consumidor.

Meios de comunicação de massa foram utilizados para fins promocionais, a fim de atingir grandes massas. A fidelidade à marca se tornou parte da cultura corporativa e novos canais de distribuição tiveram que ser criados, em que as pessoas em locais distantes teriam acesso ao fornecimento. Era importante saber quanto tempo o produto teria um significado no mercado, daí a teoria do ciclo de vida do produto. A quantidade de estudos empíricos e teóricos sobre o comportamento do consumidor cresceu e afluíram no sentido das estratégias, práticas e decisões de marketing e mercado. Este estudo, hoje,

envolve disciplinas científicas, como economia, computação, matemática, psicologia, antropologia e sociologia. O objetivo principal é detectar as necessidades e os desejos dos consumidores. É importante para o mercado incorporar as necessidades e os desejos aos produtos e serviços oferecidos (ENGEL; BLACKWELL; MINIARD, 2002).

A disciplina científica mais importante do comportamento do consumidor é a teoria psicológica, que teve início nos anos de 1960. Naquele tempo, a psicologia cognitiva estava em sua infância. Até hoje, o conhecimento ainda não está concluído, mas os resultados de pesquisas têm mostrado que, especialmente no subconsciente, desempenha grande papel nas decisões e realiza uma série de funções extremamente complexas. Os investigadores ainda descobriram que 45% das pessoas pensam sobre diversas outras coisas quando realizam uma tarefa (MARTIN, 2008).

De acordo com Kröber-Riel et al. (2008), os processos cognitivos são importantes para reconhecer como o ator ou os atores gravam informações, processam e como as armazenam. É importante a possibilidade de estimular uma pessoa e saber como ela irá reagir após a estimulação. Cientistas como Engel, Blackwell e Miniard (2002) e Solomon (2002) colocam essas informações no que denominam modelos de decisão e comportamento do consumidor.

### **2.2.1 Modelo de decisão do consumidor**

Modelos de decisão do consumidor são necessários para encontrar melhores decisões táticas e estratégias de ações e práticas de marketing e negociação. O comportamento do consumidor é um processo de ações e decisões até o consumidor escolher e consumir um produto, descartá-lo e reutilizá-lo (ENGEL; BLACKWELL; MINIARD, 2002). De acordo com Solomon (2002), é importante estudar o comportamento do consumidor antes, entre e após a

compra. Os modelos de processos de comportamento dos consumidores são afetados pelas variáveis ou fatores internos e externos (Figura 6).

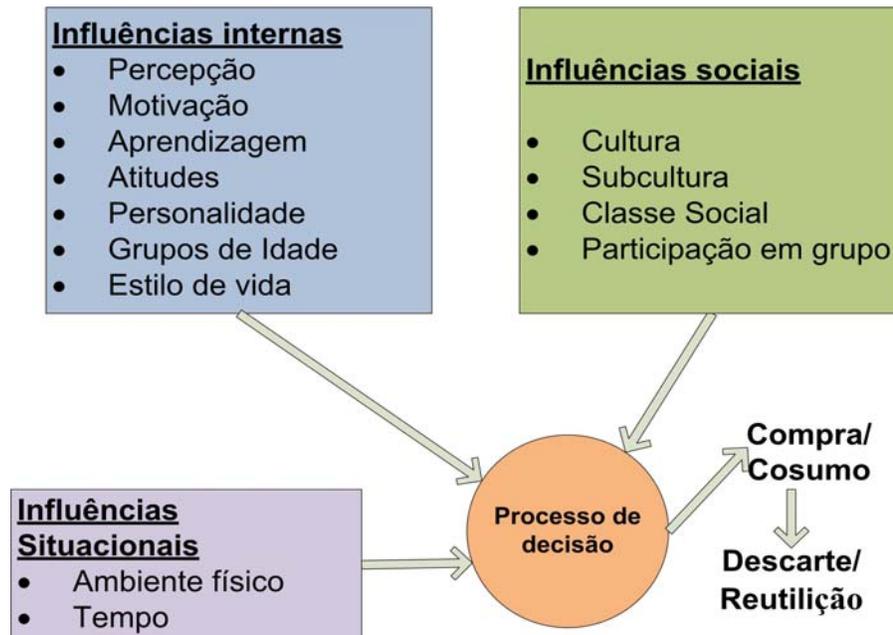


Figura 6 Fatores que influenciam o processo da decisão  
Fonte: Solomon (2002)

Os fatores internos incluem percepção, motivação, aprendizagem, atitudes, personalidade, idade e estilo de vida do consumidor. As influências do comportamento externo incluem cultura e grupos de referência e da família. As variáveis internas e externas que influenciam o processo de decisão em si são como o processo de informação do consumidor. Incluem-se também neste modelo as causas gerais e situacionais, como ambiente físico, tempo e histórias do consumidor durante o processo de consumo, de modo geral (ENGEL; BLACKWELL; MINIARD, 2002).

O estudo do comportamento do consumidor é o exame de indivíduos e grupos que mudam seu desenvolvimento e suas propriedades. Incluindo os processos de seleção, aquisição e utilização de produtos ou serviços de

indivíduos ou grupos que chamamos de teoria do papel. A teoria do papel é importante para o consumidor como indivíduo ou grupos reconhecidos no mercado. De acordo com Solomon (2002), o consumidor pode ser uma única pessoa, que terá uma necessidade ou um desejo de consumir um produto ou serviço, como também pode constituir um grupo de pessoas (compradores, usuários, influenciadores) que compram, adquirem e consomem um produto. Além disso, mesmo os papéis individuais podem ocorrer sob a forma de grupos ou organizações (SOLOMON, 2002).

#### **2.2.1.1 Influências sociais**

É muito importante explorar o comportamento dos consumidores para determinar o estímulo do mercado. Esses consumidores são especialmente importantes para lançar um produto, para anunciar a mensagem publicitária direto para o grupo alvo correto, assim significando, ao mesmo tempo, o começo do sucesso da empresa.

Solomon (2002) divide as variáveis externas ou influências ambientais, como Engel, Blackwell e Miniard (2000), em influências culturais, subculturais, classes sociais e grupos de referência. Não são apenas os grupos de referência os mais específicos. Engel dividiu os grupos de referência específicos em grupos familiares, membros da família e outros grupos. Fala-se de uma cultura comum quando um grupo de pessoas, por exemplo, vive em uma sociedade formada e dividida com pontos de vista comuns. Essas opiniões, de acordo com Solomon (2002), estão sujeitas a um sistema de valores comuns, crenças e ou uma ideia política comum. Como exemplo, tem-se a cultura alemã, o seu sistema democrático, as necessidades de trabalho, da pontualidade e da fé cristã, que são importantes bens culturais com ponto de vista comum (SOLOMON, 2002).

A cultura é uma questão amplamente estudada e importante nas áreas de pesquisa em comportamento do consumidor, de marketing e da sociologia; dificilmente, a cultura definida em um conceito unânime tem um amplo campo para tentar ser respondida e testada por outros. Segundo Schiffman e Kanuk (2000), a cultura considera as dimensões específicas no sentido de regular o comportamento humano. Os autores definem a cultura como “a soma das crenças, valores e costumes aprendidos que servem para direcionar o comportamento de consumo dos membros de determinada sociedade”. Assim, pode-se pensar que a cultura é um dos valores que caracterizam certo grupo de indivíduos, certo tipo de consumo e certo tipo de produtos a serem introduzidos nos seus hábitos do dia a dia.

O impacto da cultura em nosso meio permite criar vários segmentos de Mercado. Ao contrário de pensar que é de alguma forma algo mais restringido, acredita-se que todas essas novas manifestações de valores e crenças abrem um caminho para novas indústrias e novos produtores que talvez não tivessem o espaço para ser inseridos nos mercados.

A influência da economia mundial e as mudanças tiveram como resultado uma grande exposição e fluxo de informações. A cultura é altamente influenciada e dinâmica, e entre as mais fortes influências estão a forma de consumo, os produtos e as tendências de comportamento de países como Estados Unidos (SCHIFFMAN; KANUK, 2000). O significado da cultura reside no sistema de consumo dos Estados Unidos e a forma como este sistema é capaz de transferir esses valores para outros lugares, mesmo que as distâncias sejam consideráveis. Para o autor, a cultura depende, principalmente, de fatores materiais; cada um dos produtos que conduzem a compra tem um significado além do seu valor. Esses significados são dinâmicos e podem fazer parte de diversos tipos de avaliação em determinado tempo.

Da mesma forma, Schiffman e Kanuk (2000) afirmam que a cultura tem uma porção intrínseca ou invisível, só que, para estes autores, esse valor intrínseco não é devido a um determinado bem de consumo, mas pode ser aquelas características que são próprias de um lugar, de uma família, de uma região e pelo fato de uma pessoa pertencer a aquele lugar já ter adicionada ao seu comportamento esta característica que faz parte dessa cultura. Assim, pode-se afirmar que a cultura é um aprendizado. Nasce-se com ausência de cultura e, com o passar do tempo, os indivíduos vão sendo configurados à sociedade da qual fazem parte. A transmissão de valores, de crenças e de costumes que são aprendidos dá ao indivíduo a condição de ser membro de um grupo social.

As pesquisas feitas na área de cultura com o objetivo de analisar o comportamento do consumidor, ou fazer marketing para empresa, ou sociologicamente decifrar certo grupo social, contam, hoje em dia, com sistemas de processamento e análise que tentam medir e avaliar valores culturais para identificar presença ou ausência de valores amplamente utilizados para a análise do comportamento e dos costumes daquele grupo ou da sociedade. Segundo Schiffman e Kanuk, (2000), a análise de conteúdo e o trabalho de campo acompanhado por instrumentos de pesquisa, como programas ou sistemas que oferecem identificação e rastreamento de valores culturais e tendências sociais, podem ajudar a refletir e a interpretar comportamento em diversas culturas.

Em contrapartida, subcultura é um grupo de pessoas de uma cultura existente, com proposição de valor diferente ao da cultura-mãe. De acordo com Solomon (2002), uma subcultura caracteriza-se com base em algo em comum, como gosto de música, cinema, desporto, livros e outras formas. A subcultura não é criada por si só, de modo que o marketing influencia o seu interesse (SOLOMON, 2002).

De acordo com Solomon (2002), uma classe social representa a posição comum das pessoas dentro de uma sociedade. As pessoas que vivem na mesma

classe social têm, frequentemente, as mesmas condições de emprego e de renda. Essas pessoas têm os mesmos interesses na escolha de suas roupas ou de suas atividades. Também podem ter a mesma religião ou o mesmo interesse político.

De acordo com Engels, Blackwell e Miniard (2000), classe social é um grupo de membros que compartilham muitos valores comuns, como interesses e comportamentos, comparados em termos de ocupação, renda, educação, assistência social e outras variáveis.

Os membros do grupo, ou grupos sociais, surgem a partir de interesses comuns das pessoas, independentemente da sua classe social. Esse grupo pode ser composto por alguns membros ou mais (SOLOMON, 2002). Por exemplo, um grupo social poderá ser formado por uma comunidade ou por um amigo.

#### **2.2.1.2 Influências internas**

As variáveis internas incluem motivação do consumidor, aprendizagem, atitude, personalidade, idade e influência do seu estilo de vida individual. Essas configurações internas, de acordo com Kröber-Riel et al. (2008), direcionam a publicidade. Objetiva-se, por exemplo, prendê-lo a uma marca emocional (emoção, motivação). Esse método é utilizado também na publicidade direcionada. A teoria da aprendizagem e a teoria da percepção do consumidor são utilizadas para atingir o consumidor com a mensagem publicitária direta. As influências internas são controladas por processos mentais do consumidor. Há muitos nomes diferentes para descrever os processos mentais; grosseiramente falando, os processos de pensamentos são processos inconscientes e conscientes.

Nos últimos vinte anos, o conhecimento dos nossos processos mentais tem aumentado. O processo de pensamento subconsciente é determinado a processos de pensamentos conscientes (MARTIN, 2008).

O *design* de produtos, por exemplo, refere-se apenas à ação racional e aos hábitos ruins dos consumidores. É de grande importância para a empresa descobrir como todos os processos mentais reagem quando são confrontados com marcas, promoção e produto. O nosso subconsciente é controlado pelas emoções e elas ajudam-nos com os aspectos importantes do nosso meio ambiente, como se através de uma lente devessem ser incorporadas, a fim de filtrar e priorizar informações. Mas, o subconsciente tem uma variedade de funções vitais como batimentos cardíacos, temperatura corporal e respostas de milhares de comportamentos, tais como o corpo reage a determinadas situações (MARTIN, 2008).

As funções cognitivas não agem sozinhas, elas estão diretamente relacionadas com o processo de pensamento executivo que assume a recuperação do pensar e do armazenamento de memória e o pensamento consciente assume a lógica para resolver problemas. O cérebro executivo pode ser pensado no futuro, assim como no passado, contrastando com o processo de pensamento cognitivo que age apenas no presente, mas centenas de pensamentos podem ser processados simultaneamente. As funções cognitivas do cérebro trabalham em conjunto com as funções executivas do cérebro, mas o sentido inverso desse processo não ocorre. Assim, o pensamento racional não está conscientemente acessível para as funções cognitivas do cérebro (MARTIN, 2008).

Segundo Ash (2008), a internet é um novo meio que se tem para incluir funcionalidades que estimulam o processo cognitivo. Especialmente a página de destino deve possuir mensagens que estimulem a memória cognitiva. De acordo com Martin (2008) o cérebro recolhe informações cognitivas que são não-verbais.

A percepção é um processo que explica como o consumidor cria e interpreta a informação. Com a atenção começa a interpretação seletiva de

informações armazenáveis. A atenção deve, primeiramente, ser alcançada, a fim de dedicar-lhes a mente racional. A percepção, ou melhor, o armazenamento das informações determina o papel de uma marca, uma empresa e a importância da publicidade (MARTIN, 2008). Também é efeito do processo de consumo a maneira como o consumidor aprende. A teoria de aprendizagem é resumida pelo hábito e pelo paradigma de reforço, como o comportamento do passado nos ensina e a aprendizagem está mudando o nosso comportamento automaticamente (SOLOMON, 2002). A motivação vem de um desejo de consumo das necessidades internas. A teoria do processo de contratação de necessidades mais utilizada é a hierarquia de Maslow. A motivação é também uma dimensão dinâmica e a do consumidor pode aumentar quando ele encontra rotinas familiares em um produto.

Existem três formas diferentes para armazenar informações na memória, que são a aprendizagem visual, a aprendizagem auditiva e a sinestésica. Por exemplo, a informação visual pode ser apresentada por uma demonstração, um código de cores, diagramas, tabelas, gráficos, fotos, mapas e vídeoclipes. A aula acústica pode ser feita com a ajuda de clipes de áudio, declarações orais ou apresentações, poemas, rimas, associação de palavras, clipes de vídeo, suporte por telefone ao vivo. Já o ensino sinestésico apresenta-se por meio de atividades interativas, as emoções associadas com os conceitos como adereços ou exemplos específicos de resolução de problemas (SOLOMON, 2002).

As pesquisas mostram que a maioria das pessoas tem um estilo predominante de aprendizagem. Existem algumas que têm um equilíbrio entre duas ou três modalidades. Portanto, faz sentido o uso de um nível equilibrado de todos os métodos de aprendizagem que melhor se apela a todos os grupos.

Segundo a teoria freudiana, a personalidade são motivações e impulsos, ou seja, uma parte do nosso agir não é planejado, nem pensado, existem então uma parte involuntária que não tem como ser explicada e que, ao contrário da

ideia de uma personalidade duradoura, pode mudar, dependendo da situação. Por isso, a teoria freudiana poderia explicar como alguns comportamentos de consumo humano são inconscientes e inexplicáveis. Dessa forma, em pesquisas desenvolvidas nesta área estudam-se estes comportamentos para tentar entender e conhecer as motivações das pessoas e assim ter poder na hora de influenciar uma compra.

Conforme Monteiro e Vega (2006), compreender a relação entre consumo e personalidade é um desafio; as teorias sobre personalidade carecem de uma base empírica sólida que sustente aplicações práticas ou relações com outros construtos complexos em teorias abrangentes. Dessa forma, existem conflitos entre as diversas teorias, o que impede a análise do comportamento de uma forma homogênea para cada uma das classificações de personalidade.

Acredita-se que, dentro da nossa sociedade, existem elementos que são utilizados para tentar explicar e classificar as pessoas dentro de certo grupo; a utilização de cores, os objetos e a preferência por marcas indicam certos traços que, de alguma forma, identificam as pessoas e seus desejos a serem expressos. Vários autores determinaram, por meio de pesquisas, as diferentes motivações para desenvolver certo tipo de personalidade; para alguns, as motivações são coletivas, no sentido de consideração com as pessoas, relacionamentos sociais; para outros já são motivações mais individuais.

Segundo Coolidge et al. (2000), há três tipos de personalidade: os complacentes, os agressivos e os afastados, excluindo-se, assim, a interferência e a influência das outras pessoas. Acredita-se que tanto as motivações sociais como as individuais definem a personalidade e os valores das pessoas.

A necessidade que as pessoas têm de fazer parte de um grupo, de sentir total identificação com certos tipos de produtos e de serviços, dá um valor ao ser, o qual pode ser considerado intangível, mas que está presente em cada uma das nossas decisões de comportamento, seja ou não de consumo. O conceito de

valor está presente no nosso dia a dia, há uma diferença de significado devido ao amplo uso desta palavra e seus diversos significados. Identificam-se claramente pelo menos três conceitos amplamente aplicados a valor: o primeiro se refere a uma relação de custo e benefício; o segundo significado se refere ao valor que um cliente tem para uma organização e, finalmente, o terceiro, baseado na escola da psicologia social, aborda o aspecto do valor relativo à própria condição da existência humana em suas relações sociais, assumindo que as pessoas alcançam seus valores pessoais por meio de algumas ações ou atividades específicas, dentre as quais o consumo (MELLO; LEÃO, 2008).

Assim vê-se que nós, como seres humanos, damos valor aos objetos, situações e demais, mas também precisam sentir que são valorizados, que as escolhas, os traços e os comportamentos, ou seja, a personalidade, são importantes para as organizações nas quais estão inseridos. O impacto disso é crítico tanto para a ciência quanto para as aplicações práticas no plano de gestão de marketing das organizações. As diferentes formas como as pessoas se inserem em um determinado grupo geram um valor de nível psicológico que é, na maioria dos casos, muito mais importante que o valor em termos econômicos.

De acordo com Mello e Leão (2001), consumidores traduzem atributos de produtos em benefícios que estes produzem e que, em última instância, traduzem a orientação de seus valores pessoais. Assim, a personalidade é conceituada como um “conjunto hierarquicamente relacionado de construtos intrapsíquicos que revelam consistência no transcorrer do tempo e que se combinam com o contexto para influenciar sentimentos, pensamentos e comportamentos dos indivíduos”.

O valor entendido no contexto pessoal e psicológico do ser humano, atribuído aos benefícios e satisfações, é muito mais complexo de ser influenciado que o valor econômico, o qual é resumido como uma troca de propriedades. Para Silva, Cerchiaro e Mascetti (2008), “o consumo dos bens

vem tomando formas cada vez mais diferentes do que sua simples utilidade em si, agregando continuamente fatores de identificação e de diferenciação através de seu uso”. Dessa forma, é claro como as pessoas agregam valor pessoal a cada escolha feita e se identificam com aquele objeto, bem ou serviço. Ao redor do mundo, vê-se, hoje em dia, um consumismo de certo modo um pouco exagerado. Além disso, a luta pela identificação e inserção em algum grupo segmenta cada vez mais os mercados e exige das organizações um trato individual que gera um valor agregado aos indivíduos.

O consumidor, normalmente, tenta preservar, melhorar ou alterar a sua autoimagem pela compra de produtos ou serviços (SCHIFFMAN; KANUK, 2000). São muitos os estilos de vida hoje adotados dentro da sociedade. Um exemplo claro é a da cultura *tuning*, que é uma clara expansão da personalidade refletida em um automóvel. Esse grupo de pessoas busca personalizar um carro, não só na aparência, mas também do desempenho. Essa prática, estabelecida há mais de 40 anos nos Estados Unidos, vem ganhando cada vez mais adeptos fiéis, popularizando assim as suas crenças, comportamentos e preferências dentro do contexto de automóveis. O importante é que, embora se considerem as diferenças técnicas e mecânicas desses veículos, o valor pessoal é inatingível e impossível de quantificar.

Assim, acredita-se que o conceito de valor tem uma representação pessoal intangível, sendo impossível ser quantificados o reconhecimento e a personalidade que este objeto gera para uma pessoa e/ou no coletivo para este segmento de mercado que é acrescentado para sua satisfação e adaptação com um lugar único no mundo. Mesmo assim, essa luta pela identificação é um processo que não para, mas permanece sempre incompleto, está sempre em “processo”, sempre sendo “formado” (SILVA; CERCHIARO; MASCETTI, 2008).

Essa luta pela identificação e identidade cria segmentos de mercado cada vez mais exclusivos. O papel das estratégias de marketing é escolher aqueles grupos para tentar atender aos seus desejos ou inovar com um produto desejável que proporcione satisfação e identificação com tal segmento.

Segundo Monteiro e Veiga (2007), em seu artigo sobre o Sistema VALS-2 como método de segmentação, os valores são novamente o ponto focal para diferenciar os mercados e os objetivos das organizações. Nesta proposta, existem os valores terminais, os quais representam significados da vida e os valores instrumentais, que representam os padrões de comportamento. Como se vê, o significado de valor em comportamento do consumidor é mais complexo do que o significado de valor só em termos econômicos.

Acredita-se que um modelo de segmentação de mercado serve para tentar uma homogeneização dentro de cada grupo. Variáveis como idade, renda, nível de escolaridade e estilo de vida permitem que as organizações produzam bens e serviços para um público objetivo. A segmentação é uma ferramenta básica que identifica o indivíduo e coloca-o em um contexto em que ele se sente aceito, identificado e compatível. Além disso, é de grande potencial prático a possibilidade de as empresas segmentarem seus mercados e posicionar seus produtos segundo a estrutura de personalidade. Também reproduz um papel, o estilo de vida do consumidor.

A referência de estilo de vida não é só para o grupo de idade, mas também está relacionada à renda e aos interesses pessoais do consumidor. O estilo de vida de uma pessoa descreve como essa pessoa gasta o tempo e o dinheiro, a fim de criar valores. O estilo de vida inclui valores materiais e imateriais. É também evidente que os valores materiais não podem ser separados dos valores imateriais. Por exemplo, a decisão de comprar uma cruz com o valor da fé, não quer dizer que não vai conter o valor material da jóia (SOLOMON, 2002).

### **2.2.1.3 Influências situacionais**

O modelo de processo de consumo de Engel, Blackwell e Miniard (2000) tem as variáveis internas e externas e a influência do tempo de eliminação como fatores entre os processos de consumo. Em geral, o consumidor compara os produtos mais intensivos com menos oportunidades, disponíveis em um ambiente agradável.

Também, o tempo de consumo disponível tem um impacto sobre como o consumidor recebe a informação e avalia os produtos. Com uma grande oferta, e pouco tempo, fica muito difícil para o consumidor tomar uma decisão racional (SOLOMON, 2002).

### **2.2.1.4 Processos de decisão do consumidor**

Segundo Engel, Blackwell e Miniard (2000) e de acordo com a Figura 7, as fases do processo de decisão do consumidor são divididas em sequências, influenciadas por componentes internas e externas.

Observa-se, na Figura 7, que as fases são: necessidade, pesquisa, avaliação, compra, consumo, após consumo e descarte.

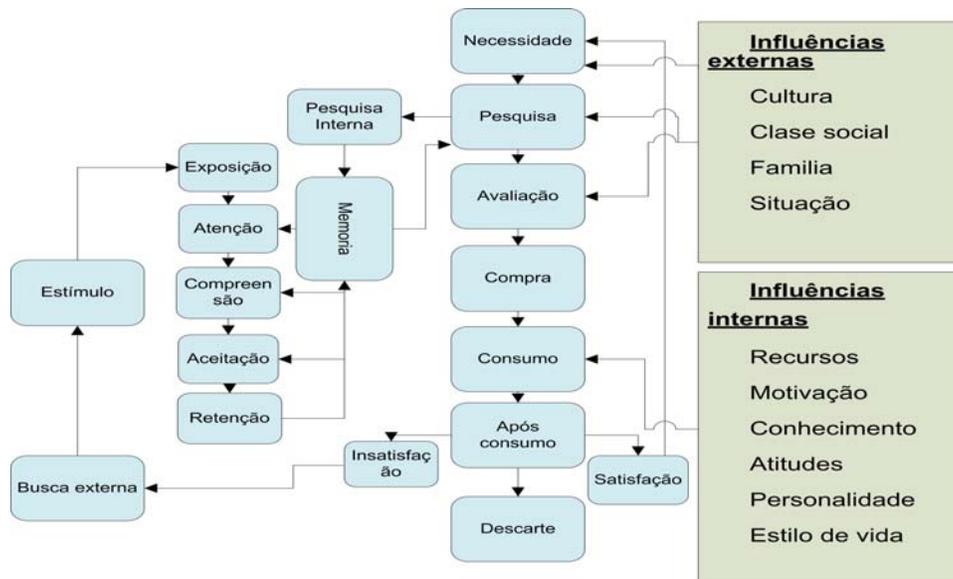


Figura 7 Modelo processos de decisão do consumidor  
 Fonte: Engel, Blackwell e Miniard (2000)

### 2.2.1.5 Necessidade

Segundo Moran (2007), os consumidores reconhecem a necessidade no princípio, sem necessariamente conhecer a solução. Para encontrar a solução, o consumidor deve pesquisar e tem essa possibilidade dividida em duas etapas (completo, insatisfatório), até que ele possa completar um novo processo de consumo. No estado completo, o consumidor está satisfeito e só começa o processo quando outra necessidade a ser buscada for definida. No estado insatisfatório, ele não decidiu e pode sofrer influências internas, como fome, sede ou influências externas, como amigos ou desejos (ENGEL; BLACKWELL; MINIARD, 2000).

### **2.2.1.6 Pesquisa - busca por informação**

Os consumidores, por exemplo, buscam informações com amigos, conhecidos, outras pessoas ou organizações. Outra fonte é o conhecimento pessoal ou a experiência pessoal. A experiência pessoal é utilizada para a pesquisa e para a avaliação da publicidade ou informação promocional (SOLOMON, 2002).

Um novo meio pelo qual o consumidor pode buscar informações é a internet. O principal objetivo, neste contexto de busca de informação, deve ser o de informar sobre produto, organização ou serviço. Antes de uma compra, o consumidor quer ter informações do produto ou serviço na Web. Quando os consumidores procuram por informações, eles podem deixar informações sobre seus interesses e isso pode ser utilizado para redesenhar uma página web (MORAN, 2007).

### **2.2.1.7 Avaliação das possibilidades**

Ao avaliar possibilidades, fatores como preço, marca, qualidade e origem dos produtos desempenham um papel fundamental na seleção de um produto (ENGEL; BLACKWELL; MINIARD, 2000). De acordo com Solomon (2002), o consumidor faz a avaliação das possibilidades, por exemplo, por meio de uma lista de prós e contras. Assim, o consumidor não necessariamente atua de forma irracional. Se ele faz uma lista de prós e contras e, depois, escolhe, deve ser racionalmente. As decisões podem ser, portanto, baseadas em critérios de avaliação que geram valores próprios, benefícios ou outros atributos. Nota-se, ainda, que alguns consumidores fazem suas decisões de compra sozinhos (ENGEL; BLACKWELL; MINIARD, 2000).

### **2.2.1.8 Compra**

Segundo Engel, Blackwell e Miniard (2000), o processo de compra pode ser influenciado, principalmente, pela pressão do tempo, pela falta de informação ou pela pressão social e psicológica. Um processo de compra pode ser diferente para a mesma pessoa em momentos diferentes.

### **2.2.1.9 Consumo e após consumo**

Quando um cliente fez uma escolha e o consumo de um produto, ele toma uma decisão sobre o produto ser bom ou ruim. Isso serve para o marketing do cliente, para definir o valor da lealdade do consumidor. Muitas vezes, as empresas taxam os clientes com questionários ou observando os padrões de compra para avaliar se eles têm ligações com a empresa ou se houve alterações no produto.

### **2.2.2 O comportamento do consumidor online**

Segundo Emerick e Round (2000), a internet é uma aplicação de cliente-servidor que fornece uma interface gráfica com o usuário para a apresentação de informação e comunicação interativa. De acordo com Wind, Mahajan e Guther (2001), o surgimento das novas tecnologias influenciou o comportamento do consumidor. Também foram observados impactos nos planos e modelos de negócios das organizações. Como resultado tem-se mudança nos contextos de marketing e estratégia empresarial orientados para o mercado (Figura 8).



Figura 8 O modelo do comportamento de um consumidor online  
Fonte: Wind, Mahajan e Guther (2001)

Assim, a internet se tornou não apenas mais uma plataforma para vender produtos, mas também uma plataforma que fornece às empresas a oportunidade de coletar dados sobre os clientes e avaliar o desempenho de um website ou uma campanha, muito rapidamente. Testes de desempenho e avaliação automática do comportamento dos clientes podem apoiar decisões de negócios e marketing (BURBY; ATCHISON, 2007).

Consumidores utilizam, frequentemente, as máquinas de busca e utilizam palavras-chave para encontrar produtos online. No entanto, nem todos os visitantes farão uma decisão de compra e a palavra-chave é apenas uma porta que nem sempre é compatível com os desejos do visitante e suas necessidades. A mineração de dados, ou *web mining*, é uma dos componentes que podem ajudar a selecionar e a cativar os visitantes em grupos de interesse (BURBY; ATCHISON, 2007).

De acordo com Song e Zahedi (2005), quatro fatores afetam o sucesso de um website: a capacidade de rede, o serviço web, a arquitetura local e o comportamento do consumidor. A meta do marketing online é trazer potenciais clientes para um site, como, por exemplo, por email, campanhas de *pop-ups* ou *banners* e, em seguida, medir e avaliar o tráfego de áreas específicas do site (HUANG, 2007). Além disso, o *web designer* desempenha um papel importante na captura de interesse do visitante. A fim de criar um site, o desenvolvedor tem que, primeiro, entender as necessidades e desejos do cliente em potencial. Todo o *design* e usabilidade podem influenciar o comportamento e as crenças do cliente (SONG; ZAHEDI, 2005).

A internet fornece uma maneira mais específica para medir o comportamento do cliente de forma mais neutra e, assim, proporciona a oportunidade de estimar rapidamente o sucesso de uma decisão de negócios. No marketing online, a proximidade com o consumidor é reduzida a uma frequência de cliques. Dessa forma, o cliente comunica suas emoções e sentimentos com cada clique. Essas emoções são importantes para reconhecer porque eles são cruciais na decisão para a compra ou não compra (BURBY; ATCHISON, 2007).

As empresas devem ter um interesse especial nas reações de um cliente. Toda ação do cliente deve levar a uma reação da empresa, com o objetivo de oferecer melhores serviços ou produtos (BURBY; ATCHISON, 2007). Na coleta de dados das empresas, são armazenadas as páginas visitadas e as rotas de navegação. Esse tipo de dados pode descrever os processos cognitivos do cliente e pode ser utilizado para aprender sobre os clientes e o sucesso das decisões empresariais. Essa não é apenas uma situação monetária, a história do comportamento de um cliente pode ajudar a aprender com o passado e a melhorar a comercialização e as decisões futuras do negócio. Para atingir estes objetivos, o site tem que ser otimizado em um processo orientado ao cliente (BURBY; ATCHISON, 2007).

Em comunicação *online*, o relacionamento começa quando o cliente potencial entra em contato com a internet. A primeira coisa que o cliente quer quando entra é uma resposta a sua pergunta, sem precisar ler a página. Quando o usuário começa a ler, precisa encontrar a informação escrita tão fácil como as informações não escritas. A navegação do site tem que ser intuitiva. Além disso, a página deve apresentar funcionalidade e não apresentar erros técnicos (EMERICK; ROUND, 2000). A segmentação do público-alvo *online* é diferente da do mundo *offline*. De acordo com Wind, Mahajan e Guthier (2001), o problema da segmentação é que nem sempre esta apresenta as necessidades de uma pessoa individualmente. A segmentação *online* não é feita exatamente como segmentação *offline* regular porque esta inclui, por exemplo, a habilidade técnica da pessoa no ambiente do usuário, podendo-se citar a necessidade de conhecimentos básicos de sistema operacional, do software navegador web (browser) ou outras tecnologias de conhecimento do cliente.

Ferramentas de mineração de dados podem apoiar esses processos e fornecer uma visão rápida do comportamento do cliente e ajudar os grupos do segmento-alvo a terem um comportamento melhor. Os modelos mentais de mineração de dados ajudam a entender esses processos e auxiliam os analistas a entenderem o cliente, bem como os trabalhadores de negócios (HEIDRICH; LIM, 2003). O analista pode utilizar uma grande variedade de padrões, a fim de obter resultados que possam ser comparados (ZAIANE, 1999). Com esta informação, a empresa pode conhecer mais sobre seus clientes e não apenas sobre quantas pessoas visitaram aquele horário.

De acordo com Lee et al. (2001), o tráfego que gera os usuários pode ser visualizado como um conjunto de cliques em linha, mostrando como os visitantes navegam em um site. Na análise dos dados, podem-se visualizar os dados coletados de visitas de um cliente com o auxílio de um algoritmo de fluxo

de cliques. Isso pode ser utilizado para explorar o sucesso de uma campanha de marketing e o desempenho de um website.

A fim de compreender os clientes *online*, a empresa precisa entender o modelo de comportamento do cliente, pelos seus estados da comunicação ou navegação (SONG; ZAHEDI, 2005). De acordo com Song e Zahedi (2005), um modelo genérico do comportamento dos clientes *online* ajuda a organização a analisar e a avaliar a compra online (Figura 9).

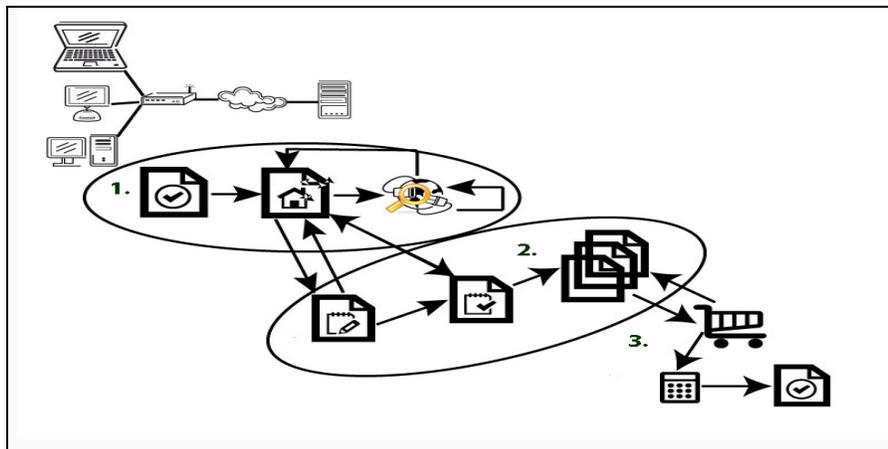


Figura 9 O processo de comunicação do comportamento do consumidor  
Fonte: Song e Zahedi (2005)

O modelo é dividido em três fases:

- a) Fase 1: o consumidor entra em uma página de destino e navega no site.
- b) Fase 2: (fase de exploração): o consumidor faz um *login* ou um registro no site.
- c) Fase 3: o consumidor faz a escolha de produtos de consumo e o pagamento.

### 3 METODOLOGIA

Neste capítulo, procura-se apresentar os componentes metodológicos utilizados na realização do trabalho. Inicialmente, são explicitados os conceitos relacionados ao tipo teórico da pesquisa, ou seja, à classificação da pesquisa. Na sequência, buscou-se apresentar os principais procedimentos metodológicos envolvendo *web mining*, considerando abordagens relacionadas a coleta, tratamento e análises estatísticas de dados.

#### 3.1 Tipo de pesquisa

Conforme Jung (2004), Marconi e Lakatos (2003), observando o método científico, tem-se que a presente pesquisa é de natureza tecnológica, com abordagem qualitativa e quantitativa, com objetivos de caráter exploratório-descritivo, utilizando procedimentos de estudo de caso e de pesquisa experimental.

A pesquisa de natureza tecnológica objetiva a aplicação de conhecimentos e tecnologias existentes, neste caso o *web mining* de uso, o estudo e o desenvolvimento de processos e serviços. Os serviços e os processos aqui relacionados referem-se à disponibilidade de um conjunto de páginas web, servindo à difusão e à matrícula em um curso oferecido ao mercado. As abordagens qualitativa e quantitativa, neste trabalho, dizem respeito, respectivamente, à busca do entendimento do relacionamento ESL de forma interpretativa, baseada nas observações e entrevista realizadas pelo investigador, e a uma interpretação numérica dos dados coletados, relativos ao cliques do usuário e à sua navegação pelo website.

O contexto exploratório descritivo se justifica por ser esta investigação fundamentada em uma análise preliminar, considerando a não existência

informações consolidadas, particularmente sobre comportamento, acessos e, mesmo, desempenho das páginas de websites educacionais. Por outro lado, buscou-se descrever efetiva e detalhadamente o processo *web mining* de uso e o comportamento do usuário no website. Procedimentos de estudo de caso e pesquisa operacional dizem respeito a: (1) delimitação de escopo do trabalho, envolvendo unicamente o website do curso de pós-graduação lato sensu a distância em Engenharia de Software Livre e (2) uso de métodos e ferramentas relacionadas à mineração de dados, funções e análises matemáticas.

Finalmente, cabe observar que, naturalmente, a investigação realizada fundamenta-se em base bibliográfica e documental, considerando contribuições científicas (publicações básicas) e documentação relacionada à universidade (UFLA), à fundação (FAEPE), à oferta e difusão do curso (marketing – propaganda) e à composição da arquitetura do website ESL.

### **3.2 Coleta e análise dos dados**

O estudo visou analisar os dados coletados, referentes às interações do usuário no website do curso de pós-graduação lato sensu a distância em Engenharia de Software Livre, ESL ([www.nte.ufla.br/esl/wp/](http://www.nte.ufla.br/esl/wp/)). Esses dados são, portanto, originários das etapas de *web mining* de uso, ou seja, do navegar do usuário no website. Eles foram coletados em dois períodos ou etapas. No primeiro período (Etapa I), trabalhou-se com um total de 1.296 usuários identificados de 1º de julho a 1º de outubro de 2009. No segundo período (Etapa II), foram 2.819 usuários, no período de 1º de janeiro a 1º de abril de 2010. Os procedimentos *web mining*, ainda a serem descritos neste capítulo, são os mesmos para as duas etapas.

Para esta coleta, foi utilizada uma ferramenta de software de apoio ao monitoramento, coleta, filtragem e tratamento, denominada Track4Mine (Figura

10). Ela registra e faz o devido tratamento de cada clique nas páginas relacionadas ao curso. Posteriormente, pode-se extrair dessa ferramenta, caso desejado, um conjunto de páginas gráficas de visualização de dados.



Figura 10 Tela da ferramenta de análise da web Track4Mine  
Fonte: Track4Mine

A ferramenta Track4Mine foi desenvolvida pelo Departamento de Ciência da Computação da Universidade Federal de Lavras (DCC-UFLA). Além dela, foi utilizado o software SPSS para algumas análises específicas. Tem-se, então, como método, a execução de todo um processo de descoberta do conhecimento em base de dados (KDD), em quatro etapas: (1) coleta, limpeza e filtragem dos dados; (2) transformação dos dados; (3), mineração de dados e (4) geração do conhecimento (análise e discussão).

### 3.2.1 Coleta, limpeza e filtragem

Com o auxílio da ferramenta de software Track4Mine são inseridas Tags Java Script (seqüência de comandos Java) nas páginas do website (Figura 11).

Conforme se observa na Figura 12, a cada clique do usuário (1) é enviada uma solicitação, no padrão web-http (<http://tmlicesa.dcc.ufla.br/>) ao microcomputador servidor da organização (2) que registra os dados no arquivo de *log*, com número de identificação-ID (3) e todas as demais informações da sessão, ou seja, do navegar do usuário. Posteriormente, o servidor retorna a informação de que registrou ao microcomputador usuário, que reinicia todo o processo (4, 5, 6 e 7).

```

Java Script of "TrackMine"
<script language="javascript" type="text/javascript"
src="http://tmlicesa.dcc.ufla.br/insyst/biclicks/coleta/Tracker.js"></script
><script language="javascript" type="text/javascript">
varTrackId=26983;CollectData();
</script>

```

Figura 11 Sequência de comandos Java Script – Tags  
Fonte: Track4Mine

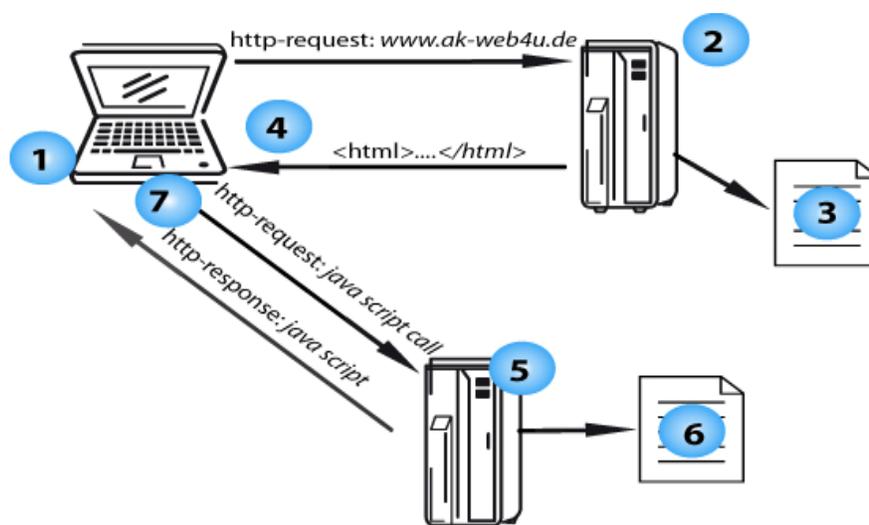


Figura 12 Base de dados – coletados e tratados  
 Fonte: Kaushik (2007)

No decorrer da coleta, os dados são limpos, filtrados e escritos em um banco de dados, sendo os irrelevantes eliminados (ZAIANE, 1999).

A cada clique, portanto, foram coletados o número do protocolo internet (IP), a resolução, o sistema de operacional, o tipo de navegador, o endereço de navegação (URL), a hora-minuto-segundo de entrada, a hora-minuto-segundo de saída, o tempo de sessão (que o usuário ficou), a página referenciada, a palavra chave utilizada para entrar na página, a cidade de origem, a região de origem e o país de origem do usuário. Todos esses dados são transformados e incluídos em banco de dados (Figura 13).

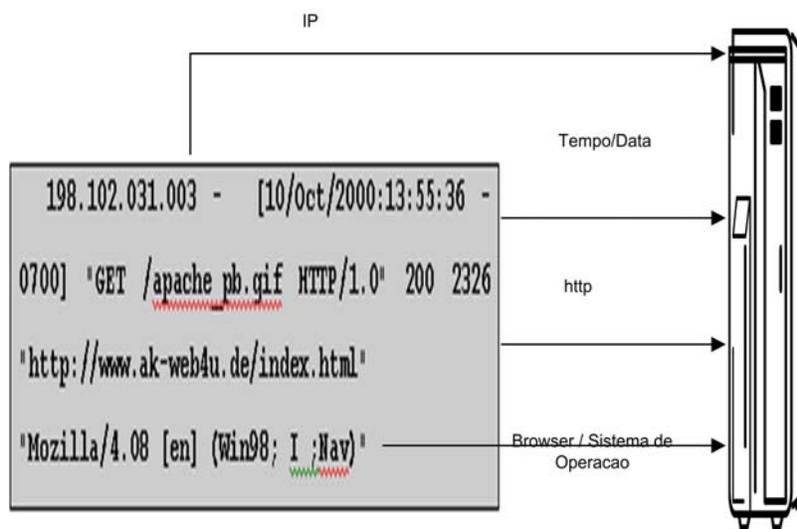


Figura 13 Envio dos dados a uma base de armazenamento  
 Fonte: Kaushik (2007)

Cada acesso à página da web é cadastrado, com todos os seus dados, na tabela principal do banco de dados. Cada dado técnico é conferido em uma tabela subordinada à tabela principal, na qual está registrado por um número de referência. Na inexistência de registro anterior, o atual é incluído nesta tabela

com um novo número de referência. Por exemplo, já existe o registro de um acesso à página da web oriundo do Brasil, então, o número de referência do país no cadastro principal de todos os acessos subsequentes oriundos do Brasil receberão o mesmo número de referência do país. Senão, será incluído um novo número de referência para o país.

### 3.2.2 Transformação dos dados

O programa Track4Mine foi o responsável pelas primeiras transformações. Cada usuário foi tratado em separado, ou seja, para cada caso, tem-se o registro de todos os cliques e passos do usuário. Com isso, podem ser replicados o caminho e o tempo. O pacote Statistical package for the social science (SPSS) foi utilizado, em alguns poucos casos, no apoio ao processo de transformação de dados.

Todas as entradas do banco de dados foram preenchidas com os dados de navegação do usuário. Assim, todo o caso tem uma sequência cronológica de páginas web armazenadas com os seus tempos associados. Além disso, foram aplicadas algumas operações, criando novas variáveis, como tempo total e médio por clique. O número de estados visitados foi resumido e armazenado como uma variável em separado.

O SPSS é um software direcionado a técnicas de estatística descritiva e temas de estatística multivariada. Por outro lado, em *data e web mining*, o fundamental refere-se a “como podemos encontrar informações escondidas (...) para decisões futuras” (NGAI; XIU; CHAU, 2009). Conforme estes mesmos autores, existem diversas e diferentes técnicas estatísticas para identificar o cliente e para explorar o comportamento do consumidor *online*.

No caso do presente trabalho, optou-se pela abordagem principal direcionada a *web mining*. Portanto, técnicas estatísticas foram utilizadas

considerando uma abordagem básica e preliminar, de análise de componentes principais e de agrupamentos, enquanto a análise de agrupamento serviu para encontrar grupos com os mesmos interesses. Esse agrupamento é chamado, no marketing, de “segmentação” e mostra grupos com mesmas atitudes e comportamento. Esses grupos expressam a segmentação tradicional e a segmentação nos grupos técnicos.

### 3.2.3 Mineração de dados

A mineração foi feita por meio de vários métodos *web mining*, em combinação com as métricas de análises da web.

Segundo Song e Zahed (2005), a fim de compreender os clientes *online*, a empresa precisa de um modelo de comportamento dos clientes na página. Esse modelo genérico do comportamento dos clientes *online* ajuda a organização a analisar e a avaliar a compra ou o fechamento do negócio *online* (Figura 14).

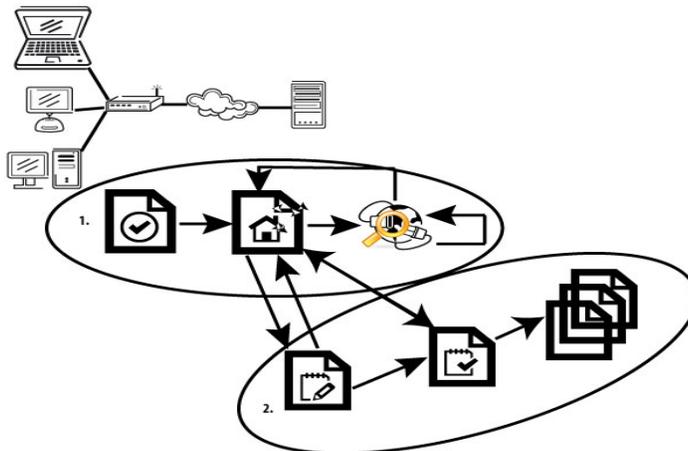


Figura 14 Comportamento do usuário nas páginas do curso ESL  
Fonte: Song e Zahedi (2005)

O modelo é dividido em duas fases: Fase 1 (fase de entrada): o consumidor entra em uma página de destino e navega no site e Fase 2 (fase de exploração): o consumidor faz uma inscrição no site e vai continuar a navegar.

Quando é feita a mineração, há a necessidade de se fazer um relatório em que o analista tem que interpretar e aprender com as análises. Esta é, talvez, a parte mais importante para análises posteriores. O responsável tem que analisar os dados e sempre avaliar porque os consumidores mudam sua atitude à medida que vão, por exemplo, adquirindo experiência com o site.

#### **3.2.4 Geração de conhecimento (análise – uso aplicado aos negócios)**

A conclusão que se pode tirar de uma análise depende dos métodos aplicados para a obtenção dos dados e da sistemática de cálculo e interpretação dos mesmos. Toda aprendizagem e decisões dependem de uma constante retroinformação (*feed-back*) dos resultados de análise obtidos até então. Com isso, a organização terá as informações necessárias para a melhoria de seu desempenho, considerando o cliente, usuário da página web.

Na fase de resultados e discussão são apresentadas as várias métricas extraídas do website, unidade caso desta pesquisa, que foram gerados com a ferramenta Track4Mine apoiada pelo SPSS. Os principais extratos de informação e métricas descritos foram: exame de página de referência, de origem do usuário, exame da taxa de retorno ou percentagem de usuários que entram e saem do website, agrupamento de variáveis tecnológicas, como número de cliques, tempo de acesso, tipo de sistema operacional utilizado, principais navegadores que os usuários utilizaram para acessar o website, estados de

origem, páginas de maior interesse do usuário, frequência da próxima fase – como buscar perceber para onde o usuário vai após determinada página, desempenho de uso de palavras na máquina de busca do Google Adwords e análises de agrupamentos específicos (clusters).

Ao final da fase de resultados e discussão tem-se, após todas as análises e estudos realizados, a apresentação de um conjunto de propostas de estratégias organizacionais, considerando o conhecimento obtido com todo o trabalho de *web mining* e as bases de informações teóricas sobre comportamento do consumidor, marketing e e-marketing.

## 4 RESULTADOS E DISCUSSÃO

A história do ensino a distância (pós-graduação lato sensu) na Universidade Federal de Lavras (UFLA) começou em 1987. Hoje, tem-se um amplo conjunto de cursos na modalidade a distância com apoio limitado de ferramentas web de negócios e ensino-aprendizagem. Os cursos são oferecidos na página web da Fundação de Apoio ao Ensino, Pesquisa e Extensão (FAEPE), parceira da UFLA, cujo endereço é [www.openufla.com.br](http://www.openufla.com.br) (Figura 15), tendo cada curso o seu próprio website.



Figura 15 Tela do website Open Ufla  
Fonte: openufla (2010)

Em 2005, o Departamento de Ciência da Computação da Universidade Federal de Lavras (DCC-UFLA) lançou o curso “Produção de Software Livre” (PSL), mas que, por sugestão de sua equipe pedagógica, após algumas ofertas,

recebeu o nome de curso de pós-graduação lato sensu a distância em Engenharia de Software Livre (ESL). O novo curso ESL iniciou-se em 2009. Diversos professores e alunos do Brasil e da América Latina participaram dele.

O website ESL constitui um conjunto de páginas que informam ao visitante sobre o conteúdo e as formas de acesso ao curso e também disponibiliza a esses visitantes, alunos e potenciais alunos o contato com a universidade por meio de um formulário específico. O website é uma página que utiliza o software de gerenciamento de conteúdo wordpress com linguagem de programação voltada para servidores Web, denominada PHP.

O website ESL inclui links para áreas ou páginas específicas (Figura 16): a primeira área (1.) inclui, além da página principal (home), também links de informações relativas ao conteúdo de curso (apresentação, metodologia e objetivos, entre outros); a segunda (2.) relaciona-se à matrícula e ao formulário de contato e a terceira (3.) aponta para links externos.



Figura 16 Tela do website do Curso de Engenharia de Software (ESL)  
Fonte: Engenharia de Software Livre - ESL (2009)

Na página inicial não podem ser vistos alguns links de segundo nível que não foram monitorados nas ações de *web mining* deste trabalho. Todas as áreas de links (principais) que aparecem em cada página em que o visitante pode clicar são monitoradas durante a visita. O usuário pode deixar o site a qualquer momento diretamente ou via acesso a links externos.

#### **4.1 A arquitetura do site ESL**

A arquitetura do website ESL é, basicamente, definida pelos três links principais especificados na Figura 16. Assim, foram considerados para *web mining* os links de primeiro e alguns de segundo níveis.

Na Figura 17 tem-se a ilustração da arquitetura em árvore das páginas considerando seus respectivos níveis. Todos os links de primeiro nível (curso, metodologia, propaganda da inscrição, objeto, disciplina, professor e credenciamento) são denominados estágios e conectados com os links do segundo nível (formulário de matrícula, confirmação). Cabe observar que as páginas formulário de matrícula e confirmação se encontram em modo tracejado, pelo fato de estas duas páginas não se encontrarem disponíveis para *web mining* na Etapa I desta pesquisa.

Em cada página encontra-se um subprograma Java Script tracker, ou seja, uma sequência de comandos em linguagem Java para coleta de dados de navegação do usuário nas páginas do website. Este subprograma envia dados para o software Track4mine, como número IP (protocolo internet), sessão (página visitada), tempo de permanência na página, data, nome do navegador utilizado e outras variáveis técnicas.

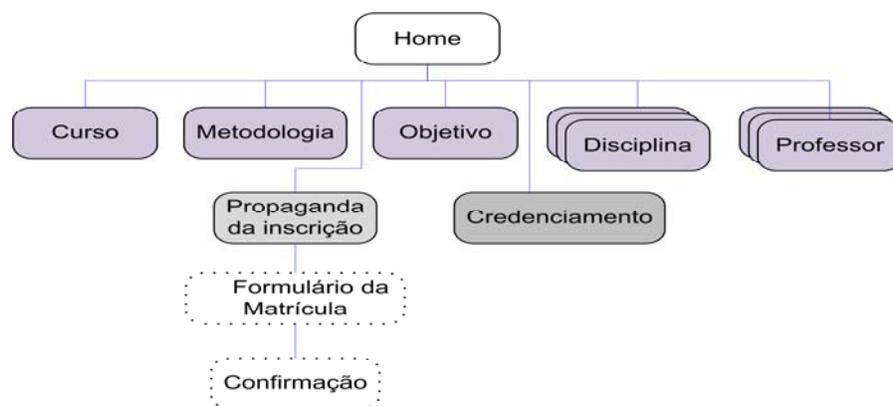


Figura 17 Arquitetura em árvore do website ESL

Fonte: ESL (2009)

No Quadro 1, tem-se a apresentação detalhada desses links, com nome da página, nível hierárquico na árvore, número da página (gerado pelo software Track4web) e descrição da funcionalidade da página. São estas páginas que serão analisadas considerando o *web mining* de uso.

Quadro 1 Descrição das páginas do ambiente do website ESL

Nome da Página	Nível	Número Página	Descrição
Home	1	P1	A página inicial
Curso	1	P2	Descrição sobre o curso
Metodologia	1	P3	Explica a disciplina Metodologia
Objetivo	1	P4	Objetivo do curso
Disciplina	1	P5	Disciplinas do curso
Curso 01	2	P5.1	Disciplina 1 "A introdução ao curso"
Curso 02	2	P5.2	Disciplina 2 "A qualidade do software"
Curso 03	2	P5.3	Disciplina 3 "Engenharia de software 1"
Curso 04	2	P5.4	Disciplina 4 "Engenharia de software 2"
Curso 05	2	P5.5	Disciplina 5 "Desenvolvimento software livre"
Curso 06	2	P5.6	Disciplina 6 "Open source software"
Curso 07	2	P5.7	Disciplina 7 "Engenharia de software 3"
Curso 08	2	P5.8	Disciplina 8 "Empreendedorismo"
Curso 09	2	P5.9	Disciplina 9 "Monografia de conclusão"
Professor	1	P6	Professores do curso
Professor 1	2	P6.1	Descrição professor disciplina 1
Professor 2	2	P6.2	Descrição professor disciplina 2

"Continua"...

„Quadro 1, continuação”

Nome da Página	Nível	Número Página	Descrição
Professor 3	2	P6.3	Descrição professor disciplina 3
Professor 4	2	P6.4	Descrição professor disciplina 4
Professor 5	2	P6.5	Descrição professor disciplina 5
Professor 6	2	P6.6	Descrição professor disciplina 6
Professor 7	2	P6.7	Descrição professor disciplina 7
Professor 8	2	P6.8	Descrição professor disciplina 8
Professor 9	2	P6.9	Descrição professor disciplina 9
Professor 10	2	P6.10	Descrição professor disciplina 10
Propaganda Matrícula	1	P7	Link para o formulário de matrícula
Contato	1	P8	Formulário de contato
Credenciamento	1	P9	Informações sobre a certificação do curso
Formulário Matrícula	1	P10	Formulário de matrícula – preenchimento
Confirmação	1	P11	Submeter o formulário de matrícula

Fonte: Dados da pesquisa

Cabe apontar novamente uma falha observada na arquitetura do website, na Etapa I, que é o não apontamento e acompanhamento das páginas formulário de matrícula e confirmação. Este erro diz respeito à necessidade de estudos detalhados e questionamentos de projeto quando da definição da arquitetura de um website, ou seja, um site projetado adequadamente permite avaliar melhor os movimentos dos usuários, o que não foi o caso do website investigado. A adequada arquitetura permite ao analista observar e registrar a sequência de movimentação do usuário no website. Essa movimentação permite, por um lado, a avaliação da facilidade ou da dificuldade que o usuário encontra em relação às funções do “menu” e, por outro, em qual sequência ele acessa as diversas páginas, correspondentes aos assuntos de interesse. Isso permite reproduzir a sequência da tomada de decisões desde o início até a saída do website.

O que se tem, portanto, é uma alternância de visitas e registros (Figura 18) que efetivamente irá viabilizar o acompanhamento do usuário no website, considerando aspectos relacionados a melhorias de desempenho de uso e potencialização de ações de marketing e relacionamento com o consumidor.

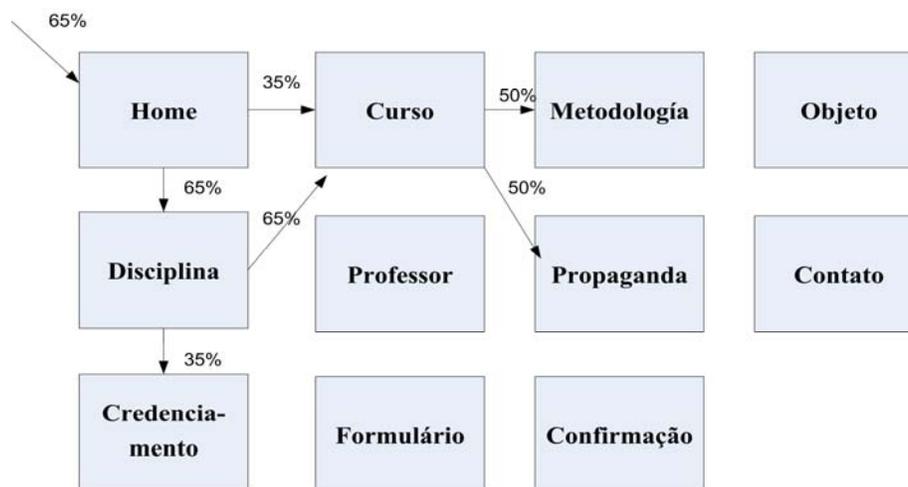


Figura 18 Exemplo de acompanhamento de visitas no website ESL  
 Fonte: Ash (2008)

#### 4.2 Divulgação e comportamento do consumidor ESL

O setor responsável pela área de marketing (divulgação) de todos os cursos da FAEPE-UFLA, conseqüentemente pelo ESL trabalha com diversas formas de publicidade, visando informar aos potenciais clientes sobre os cursos oferecidos.

Os principais veículos de comunicação e elementos de divulgação utilizados, considerando a abordagem *offline*, são jornais, revistas, TV, rádio e cartões de mídia impressa. A divulgação *online* é orientada a internet, considerando o website [www.openufla.com.br](http://www.openufla.com.br) e os websites de cada curso. Além disso, cada curso tem uma campanha fundamentada no Google Adwords (Figura 19).

O Google Adwords é uma forma de marketing na qual se utilizam palavras-chave especiais para alcançar o consumidor quando ele realiza buscas

no website da Google. Cabe observar que, no contexto do e-marketing, esta é a única atividade efetivamente realizada pela organização.

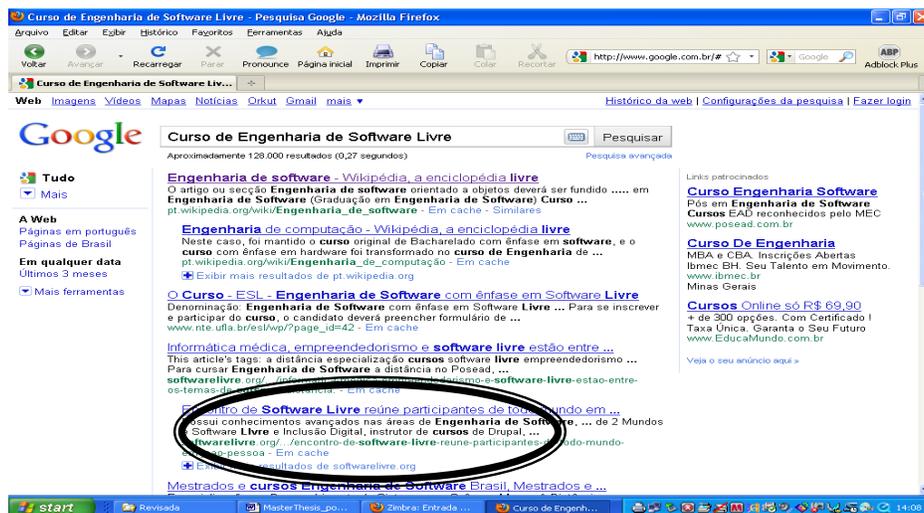


Figura 19 Busca direta do Curso ESL na web – Google  
Fonte: ESL (2009)

Para o curso de ESL, a equipe de marketing utilizou as seguintes palavras na contratação do Google Adwords: computação UFLA, curso de engenharia de software, engenharia de software, engenharia de software experimental, engenharia de software requisitos, engenharia de software uml, graduação engenharia de software, mestrado ciência da computação, mestrado computação, mestrado computação, mestrado sistemas de informação, pós-graduação computação, pós-graduação em informática, pós-graduação de engenharia de software, pós-graduação informática, pós-graduação informática e produção de software.

A grande maioria dessas palavras se mostrou ineficiente para se chegar ao curso, como por exemplo, computação UFLA (Figura 20).

Nas etapas de estudos de *web mining* que se seguem, são apresentadas as observações e as discussões, também neste contexto, ou seja, o de uso do

Google e da necessidade de utilização de palavras adequadas a um bom direcionamento do usuário, potencial aluno do curso.

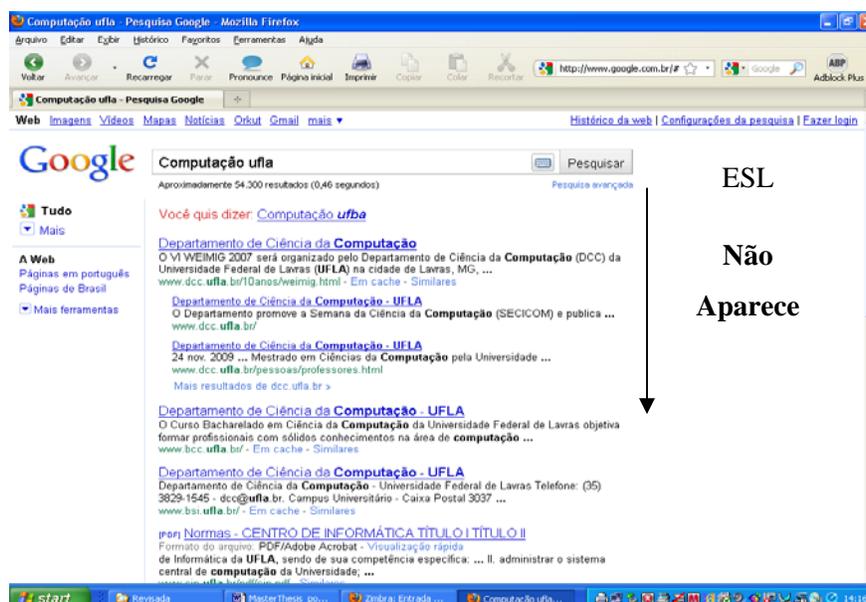


Figura 20 Busca por palavras no curso ESL no Google – Ineficiente  
Fonte: Google (2010)

O efetivo relacionamento *online*, portanto, está concentrado nos alunos que visitam os websites da organização ([www.openufla.com.br](http://www.openufla.com.br)) e do curso ESL ([www.nte.ufla.br/esl/wp](http://www.nte.ufla.br/esl/wp)). Para cada aluno que preenche o formulário de inscrição, segundo a entrevista realizada com o diretor de marketing e ensino a distância, é encaminhado um e-mail e uma correspondência (correios) contendo um “folder” com informações detalhadas sobre os cursos. Além disso, o endereço completo do visitante é cadastrado na base de mala direta, para contatos atuais e futuros.

A partir da entrevista e dos dados documentais levantados, tais como folder, e-mails, cartões de mídia e planos de ensino dos cursos, um resultado

preliminar já encontrado nesta pesquisa é que a base de marketing da organização é limitada em seu contexto de inteligência competitiva (externo) e ainda mais no seu contexto de inteligência organizacional (interno). Enfim, não se têm estudos externos ou internos que visem potencializar o processo de e-marketing e ou do e-relacionamento. Cabe observar, também, que o website ESL tem poucos elementos gráficos atrativos e expressos em suas páginas ou sequer encaminha, imediatamente, uma mensagem ao usuário, quando de um acesso ou dúvida. Também não há segmentação de usuários, tem-se uma massa de informações. Essa massa não atende, conforme já observado, a qualquer segmentação.

A “estratégia de massa” da organização, em geral, segundo as informações coletadas, é de alto custo, envolve um grande número de pessoas e material com elaboração complexa. Especificamente no caso do curso e do website ESL, não se tem nenhuma abordagem direcionada. Ocorre propaganda generalizada para todos os cursos, totalmente padronizada. Entende-se que a organização deve trabalhar um pouco mais a dinâmica da abordagem *online*, conectando relações não somente entre as páginas de um website, mas entre as abordagens *online* e *offline* – a dinâmica deve relacionar usuário visitante com elementos de propaganda de cada curso, encaminhada a este usuário interessado.

#### **4.3 Web mining ESL – Etapa I**

Inicialmente, têm-se a avaliação e a discussão da campanha que a organização realizou utilizando o Google Adwords, serviço de anúncios na forma de links, baseados em palavras-chave, do Google.

Na sequência, considera-se a primeira etapa da investigação do website (01/07 a 01/10/2009). Buscou-se expor os resultados obtidos a partir da análise de mineração de dados do ESL, da entrada do usuário na página até a sua saída,

considerando, basicamente, as métricas descritas no referencial teórico, tais como exame da campanha Google Adwords, exame da página de referência, páginas de destino e agrupamento das variáveis. Assim sendo, nesta primeira etapa, foram avaliadas as entradas e a navegação nas páginas home, curso, objetivo, disciplina, professor, propaganda da inscrição e credenciamento (Figura 21).

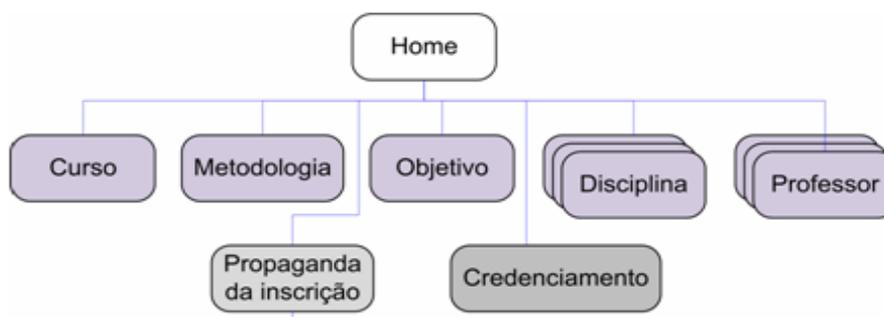


Figura 21 Páginas efetivamente monitoradas do website ESL  
Fonte: ESL (2009)

Um ponto crítico considerado nesta primeira etapa foi a ausência do monitoramento da página formulário de matrícula, ou seja, esta página não foi registrada como elemento de *tracking*. Neste caso, não se pode completar a tarefa de análise do comportamento do usuário, uma vez que não se tem ideia se ele finalizou ou não uma operação e enviou o cadastro.

Outro ponto fraco nesta primeira etapa foi a não possibilidade de análise do retorno da propaganda *offline*, quando se considera a investigação sobre acesso via [www.openufla.com.br](http://www.openufla.com.br). Este endereço também não pode sofrer uma análise *web mining*, pois foi tratado como link externo sem análise de log. O *web mining* somente se inicia quando o usuário encontra o link do website ESL.

### 4.3.1 Exame da campanha publicitária do Google Adwords

Durante o exame dos dados da campanha publicitária Google Adwords, observou-se (Figura 22) que menos de 10% dos visitantes do website entram no curso ESL utilizam uma máquina de busca.

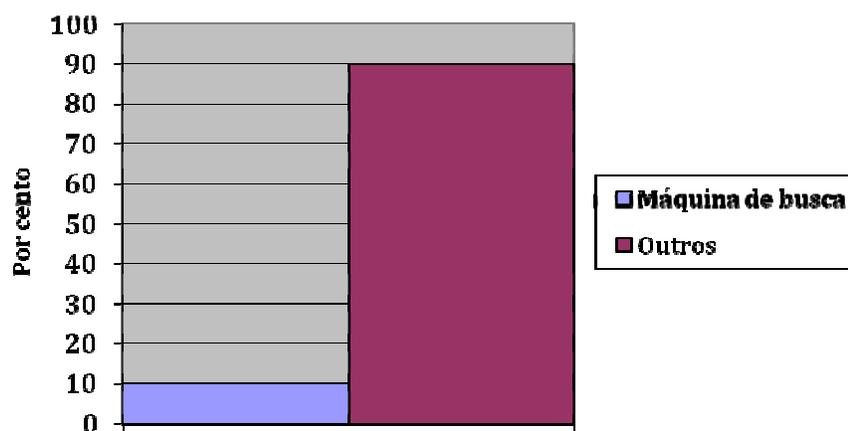


Gráfico 1 Uso da máquina de busca – Google

Fonte: Track4Mine

Nesse sentido, tem-se que a campanha no Google Adwords, com as palavras definidas pela organização (Quadro 3), não apresentou retorno. Neste quadro, quando são utilizados "---", isso significa que a palavra-chave nunca foi encontrada (usada por visitantes). Enfim, durante o tempo de medição, nunca foram efetivamente utilizadas as palavras-chave selecionadas para uso no Google AddWords.

Quadro 2 Palavras-chave Adwords usadas para acesso ao curso ESL

<b>Palavras-chave</b>	<b>Número da variável</b>	<b>Número de uso</b>
computação ufla	1110	0
curso de engenharia de software	---	---
engenharia de software	---	---
engenharia de software experimental	---	---
engenharia de software requisitos	---	---
engenharia de software uml	---	---
graduação engenharia de software	768	0
mestrado ciência da computação	---	---
mestrado computação	659	0
mestrado computação	1265	---
mestrado sistemas de informação	---	---
pós-graduação computação	1059	0
pós-graduação em informática	---	---
pós-graduação de engenharia de software	---	---
pós-graduação informática	---	---
pós-graduação informática	---	---
produção de software	---	---

Fonte: Track4Mine

As palavras-chave listadas no Quadro 4 são resultados da pesquisa orgânica (geradas pelo próprio mecanismo de busca do Google e não a partir de registro da organização). Com estas palavras, o usuário encontrou a página da UFLA e ela mostra em que palavras os usuários pensam quando procuram a página da internet.

Quadro 3 Palavras-chaves utilizadas na máquina de busca – modo orgânico.

<b>Palavra-chave</b>	<b>Número de uso</b>
Nenhuma palavra-chave	1183
Software livre	11
Engenharia de software	8
Mestrado em informática	3
Ufla engenharia de software	3
Ufla esl	3
Outros	96

Fonte: Track4Mine

### 4.3.2 Exame da página de referência

A origem de entrada do usuário na página do website ESL é variável, ou seja, ele pode vir de qualquer página, denominada de página de referência. Assim, os visitantes podem entrar, direta ou indiretamente, no ambiente do site ESL. Os links de entradas podem ser, por exemplo, uma máquina de busca, e-mail, websites de redes sociais ou quaisquer outras páginas.

Na avaliação deste trabalho, foram cinco os principais elementos de referência: (1) acesso direto, (2) via domínio da UFLA, (3) máquina de busca, (4) rede social e (5) e-mail, conforme ilustrado na Figura 23.

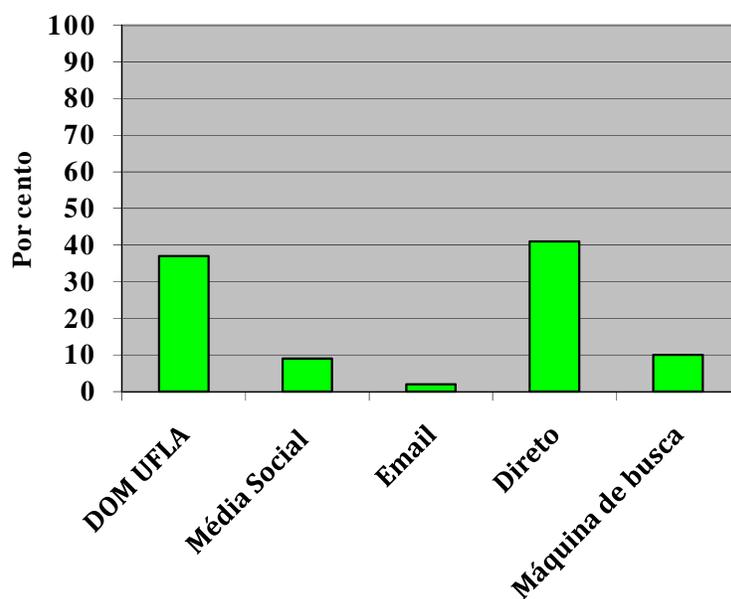


Gráfico 2 Gráfico das referências de entrada no website ESL  
Fonte: Track4Mine

Do grupo de máquina de busca (*search engine*) os usuários são

provenientes da página Google (<http://www.google.com.br>) ou da página de busca especial, oferecida também pela Google, ou seja, o URL <http://www.images.google.com.br>. O grupo referencial menor é o originário de e-mail, particularmente das URLs Google: <http://mail.google.com> e Yahoo: <http://br.mg2.mail.yahoo.com>. Finalmente, observou-se que a grande maioria dos usuários é originária dos sites da universidade ([www.ufla.br](http://www.ufla.br)) ou acessa diretamente o endereço ESL ([www.nte.esl.ufla.br](http://www.nte.esl.ufla.br)).

### 4.3.3 Exame taxa de retorno (bounce rate)

A medição da taxa de retorno (Tabela 1) expressa a porcentagem de visitantes que deixaram a página após o primeiro clique.

Tabela 1 Taxa de retorno - Etapa I

ETAPA	ENTROU	UM (1) CLIQUE	TAXA DE RETORNO (%)
Home	810	473	58,40%
Curso	038	020	54,60%
Objeto	003	001	66,66%
Disciplina	042	021	50,00%
Metodologia	011	007	63,64%
Propaganda	360	177	49,16%
Professor	009	003	33,34%
Credenciamento	002	001	50,00%
Contato	0020	016	80,00%

Fonte: Track4Mine

Esta análise é utilizada pelo marketing da organização para a avaliação do valor de informação do website. Permite detectar se a propaganda inserida combina com o site e o valor que o usuário dá ao seu conteúdo.

Se o valor for igual a 100%, indica que todos os usuários deixaram a página após o primeiro clique. Em linhas gerais, considerando critérios de negócios especificados por empresas da área de busca, conforme <http://www.mundoseo.com.br>, pode-se tomar por base o seguinte:

- a) **21%-50%:** resultados obtidos pela ampla maioria considerados normais. É possível melhorar o *bounce rate* adotando técnicas de otimização, além de focar em seu nicho específico;
- b) **51%-100%:** estes resultados são extremamente desfavoráveis. Sua taxa de retorno está muito elevada; os visitantes estão repudiando o que está sendo apresentado.

A taxa de retorno é mais indicada para a avaliação de uma propaganda *online* e sua respectiva website inicial. Uma vez que a única campanha de propaganda *online* não funcionou, pode se ter uma ideia sobre quais das páginas os visitantes deixam de imediato. Verificou-se que as páginas “professor” e “propaganda” são as mais bem colocadas, mas deve-se considerar o baixo número de acessos.

#### **4.3.4 Agrupamento das variáveis do usuário**

Nesta métrica, procurou-se classificar e agrupar alguns elementos de uso da web pelo visitante, tais como soma dos cliques, tempo gasto e resolução do navegador. Dessa maneira, foram identificadas semelhanças de comportamento de visitantes, o que permite a sua classificação em segmentos. O usuário do site tem certo número de visitas, ou melhor, de frequência de cliques na navegação pelo “menu” do site (Figura 24). Dessa forma, cada visitante executa, no mínimo, um ou mais cliques por visita. A frequência dos cliques permite calcular o percentual total da *bounce rate* pela divisão do número de visitantes de um clique pelo total de visitantes multiplicado por 100.

No caso em que após mudança das informações e ou do conteúdo do site houver queda na frequência dos cliques, fica claro que o novo conteúdo é de menor atratividade ou valor para o visitante.

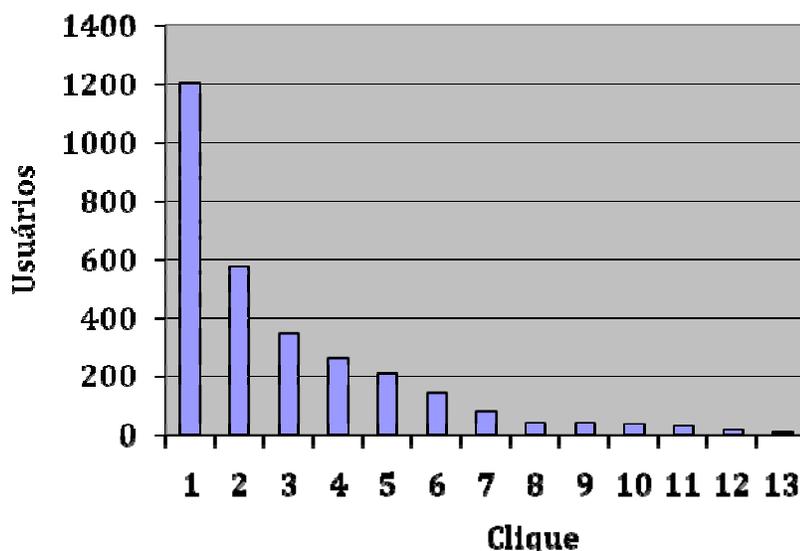


Gráfico 3 Agrupamento de cliques – Etapa I  
Fonte: Track4Mine

Outra forma de segmentação é a classificação dos usuários do site em grupos, demonstrando o tempo total da visita por colunas escuras e o tempo médio do site em colunas claras (Figura 25). No caso de um tempo médio ou total da visita menor do que 20 segundos, pode-se afirmar que o visitante pouco se dedicou às informações escritas ou à leitura do conteúdo do site. Se, após uma mudança do conteúdo do site, os tempos apurados mudam, pode-se aferir se houve melhora no tempo dedicado. Então, houve maior interesse e dedicação, ou vice-versa.

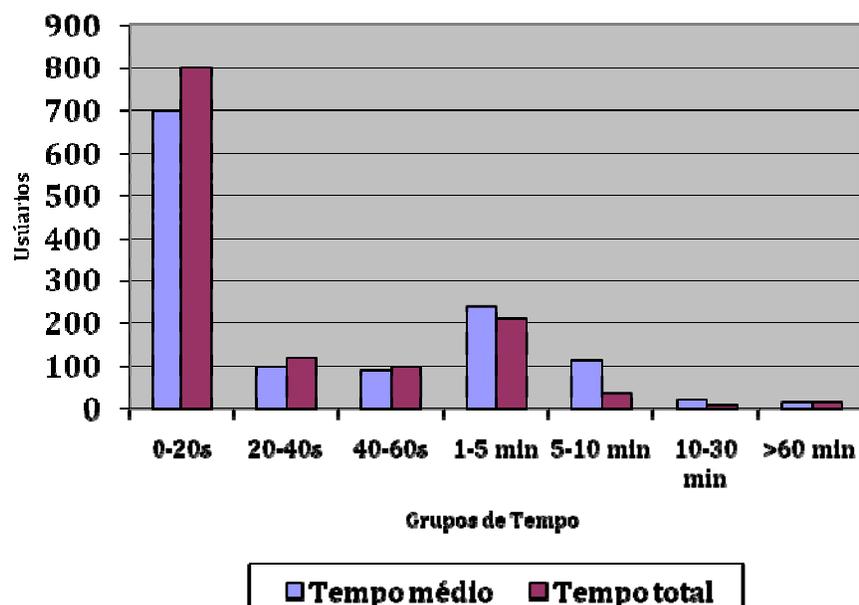


Gráfico 4 Agrupamento por tempo – Etapa I

Fonte: Track4Mine

Para se ter uma visão do conhecimento e do conjunto de ferramentas tecnológicas utilizadas pelo usuário, os dados permitem levantar: (1) o sistema operacional, (2) o tipo de navegador, (3) a resolução e (4) a segmentação demográfica. Os sistemas operacionais preferidos pelos visitantes são Windows Vista e Windows XP (Figura 26). Há também usuários que utilizam os sistemas WinServer, Win2000 e MacOS. Com relação aos navegadores, os mais utilizados são Mozilla Firefox e o Internet Explorer (Figura 27). A resolução mais utilizada (Figura 28) é 1220 x 1024 pixels e 800 x 768 pixels. Esta análise é útil principalmente para testes de atualização, uma vez que se pode redesenhar o website visando otimização para as principais preferências.

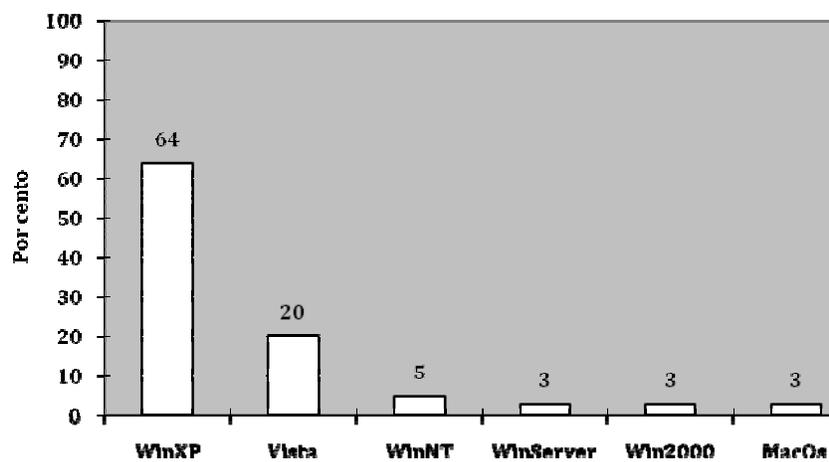


Gráfico 5 Sistemas operacionais utilizados para acesso a página ESL  
Fonte: Track4Mine

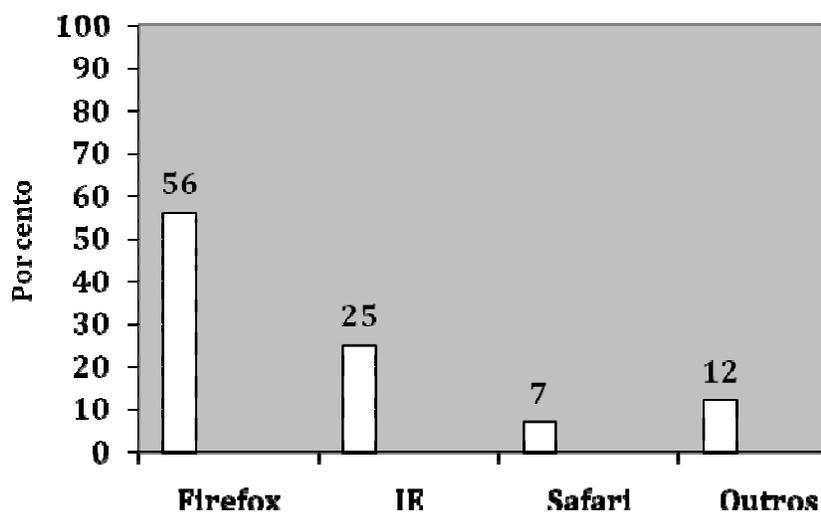


Gráfico 6 Navegadores utilizados para acessar a página ESL  
Fonte: Track4Mine

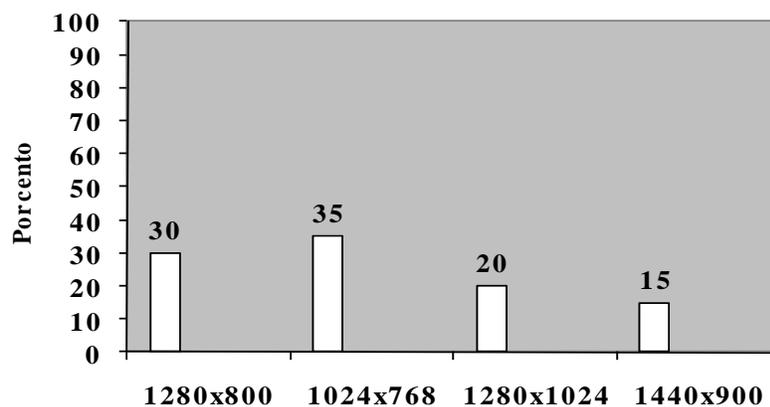


Gráfico 7 Resoluções mais utilizadas para acessar a página ESL  
Fonte: Track4Mine

O estudo da segmentação demográfica apresentado na Figura 29 ilustra que a maioria dos visitantes da página é do Brasil. Grande parte encontra-se em São Paulo, Distrito Federal, Paraíba, Rio de Janeiro e Minas Gerais.

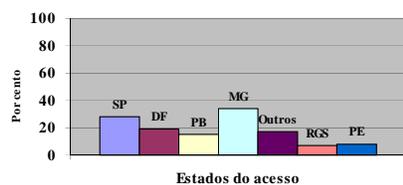


Gráfico 8 Estados de origem dos visitantes da página ESL  
Fonte: Track4Mine

### 4.3.5 Exame da atividade

A taxa de retorno não deixa claro se os visitantes estão interessados ou não no conteúdo de uma determinada página. Com a taxa de atividade, o investigador pode observar, mais especificamente, por qual página as pessoas têm mais interesse.

Tabela 2 Exame da atividade - Etapa I

	Números de páginas								
	P1	P2	P3	P4	P5	P6	P7	P8	P9
<b>Clique 1</b>	810	038	011	003	048	020	360	004	002
<b>Clique 2</b>	050	189	020	021	147	027	086	025	012
<b>Clique 3</b>	024	038	076	033	100	032	017	025	004
<b>Clique 4</b>	017	24	033	047	083	036	011	011	004
<b>Clique 5</b>	028	024	009	014	083	039	007	003	002
<b>Clique 6</b>	018	015	009	007	051	030	008	003	005
<b>Clique 7</b>	012	011	004	005	020	024	010	002	005
<b>Clique 8</b>	008	005	003	004	021	005	004	001	006
<b>Clique 9</b>	003	002	007	002	020	011	004	001	001
<b>Clique 10</b>	004	002	002	003	018	007	001	001	004
<b>Clique 11</b>	004	002	001	002	013	006	000	002	000
<b>Clique 12</b>	003	001	000	001	007	008	000	000	000
<b>Clique 13</b>	000	001	001	000	004	005	000	000	000

Fonte: Track4Mine

Por exemplo, na Tabela 2 mostram-se os primeiros 13 cliques de todos os casos. Observa-se que a concentração de cliques se encontra na página home (P1) e na página propaganda da matrícula (P7). Do segundo até o sétimo clique, a maioria das pessoas está visitando a página curso (P2) e a página disciplina (P5). Nos próximos cliques, os visitantes querem informações sobre a página disciplina (P5) e professor (P6). Menos interesse eles tiveram pela página credenciamento (P9).

Na Tabela 3 têm-se os dados de taxa de atividade (*trough rate*). Esta tabela refere-se ao resultado da Tabela 2. O maior impacto tem a página home

(P1), depois a página disciplina (P5) e a página propaganda da matrícula (P6). Parece um pouco diferente porque os primeiros cliques são contados do clique da taxa de atividade. Mas também é muito claro que os visitantes têm muito interesse na página disciplina (P5) e pouco na página credenciamento (P9).

Tabela 3 Exame da atividade - Etapa I

Fase	Cliques	Exame da atividade
Home	981	29,83
Curso	352	10,70
Metodologia	176	5,35
Objeto	142	4,32
Disciplina	757	23,01
Professor	250	7,60
Propaganda da matrícula	508	15,45
Contato	78	2,37
Credenciamento	45	1,37
Impressos	3289	100

Fonte: Track4Mine

#### 4.3.6 Frequência da próxima fase

A navegação pelo site também pode ser analisada por meio de regras de associação. Cada usuário tem caminhos de cliques que são transferidos para uma matriz. No Quadro 5 pode-se observar que o Usuário 4 entrou pela página de propaganda de matrícula (P7), em seguida passou pela sequência: curso (P2)=>metodologia (P3)=>objetivo (P4) =>disciplina (P5) =>professor (P6).

Quadro 4 Exemplo da intercomunicação da população nesta pesquisa

USUÁRIO	PN1	PN2	PN3	PN4	PN5	PN6
Usuário 4	P7	P2	P3	P4	P5	P6
Usuário 5	P1	P2	P1	P2	P7	P2

É possível, a partir desta navegação, tentar prever que página o usuário irá acessar no próximo clique. Cada atividade do usuário é guardada em uma

matriz, como mostrado na Tabela 4. O usuário 5 vem de fora (start) e entra na página inicial home (P1) (coluna), adicionando 1 para a coluna P1 em linha iniciar. Então, ele vai de home (P1) para curso (P3), adicionando 1 para a coluna curso (P3) na linha home e assim por diante, até que o caminho do clicar seja acabado. Todos os números das páginas são listados no capítulo 4, seção 1, arquitetura do site ESL. Quando todos os caminhos do usuário são salvos na matriz, cada célula é dividida pela sua soma total de linha, resultando em porcentual ou possibilidade de o usuário clicar em um link (coluna), quando estiver em uma página específica (linha).

Tabela 4 Matriz de frequência na próxima fase - Etapa I

PAG.	Iniciar	P1	P2	P3	P4	P5	P6	P7	P8	Fim
<b>Iniciar</b>	0	0,65	0,02	0,01	0,04	0	0	0,27	0	0
<b>P1</b>	0	0,12	0,32	0,05	0,2	0,03	0,02	0,19	0,03	0,03
<b>P2</b>	0	0,18	0,02	0,39	0,18	0,03	0,01	0,08	0,09	0,02
<b>P3</b>	0	0,08	0,16	0,01	0,19	0,04	0,01	0,04	0,46	0,02
<b>P4</b>	0	0,1	0,08	0,04	0,06	0,18	0	0,07	0,06	0,42
<b>P5</b>	0	0,21	0,08	0,04	0,16	0,01	0,01	0,01	0,04	0,45
<b>P6</b>	0	0,2	0,18	0,03	0,07	0,13	0,02	0,17	0,05	0,15
<b>P7</b>	0	0,13	0,31	0,03	0,19	0,02	0,09	0,18	0,01	0,04
<b>P8</b>	0	0,08	0,06	0,12	0,6	0,08	0	0,02	0	0,03
<b>Fim</b>	0	0	0	0	0	0	0	0	0	0

Fonte: Track4Mine

Observa-se que a maioria das pessoas entra na página home (P1) e na página propaganda da matrícula (P7). Todos os números são listados na descrição da arquitetura. Os números das linhas mostram onde o usuário pode ir quando do próximo clique. Com isso, pode-se pensar no caminho principal: P1 (home)->P2 (curso)->P3 (metotologia)->P8 (contato)->P4 (objetivos)->P9 (fim); ou P7 (promoção de matrícula)->P2 (curso)->P3 (metotologia)->P8 (contato)->P4 (objetivos)->P9 (fim). O usuário está saindo principalmente depois deste caminho. Se ele visitar a página de contato, isso pode significar que ele está procurando mais informações, indisponíveis nas páginas visitadas.

#### 4.4 Webmining ESL – Etapa II

Na pesquisa realizada considerando a Etapa I, o maior problema foi que a página formulário de matrícula não se encontrava integrada ao *web mining*. Na realidade, um grande erro de projeto e estratégia de propaganda. A matrícula é a finalidade principal, ou melhor, o destino real do website ESL. Nesta nova medição, a página formulário de matrícula foi devidamente integrada à arquitetura do website (Figura 31). Além disso, criou-se um mecanismo para identificar se o usuário enviou o formulário.

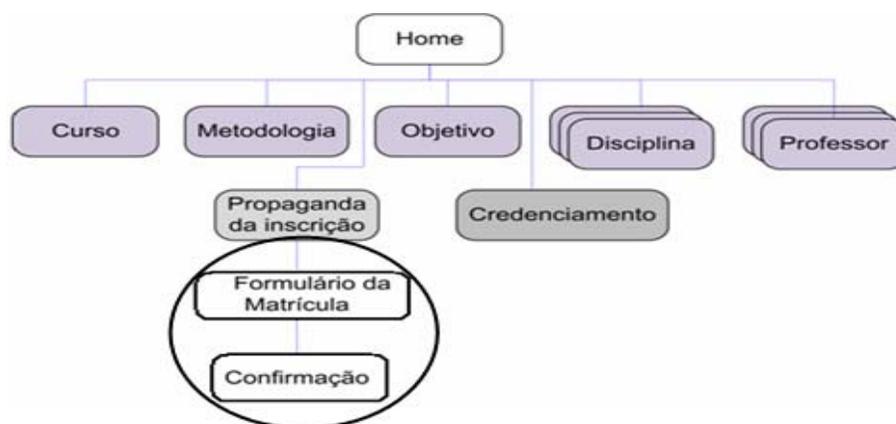


Figura 22 Arquitetura do website com formulário de matrícula e confirmação  
Fonte: ESL (2009)

Com esta nova página tem-se um funil de conversão, encaminhando o usuário para este destino, visando alcançar um dos principais objetivos do website e, mesmo, da organização. Também agora serão novamente analisados primeiros os links de entradas e comunicação dos usuários até que eles efetuem uma matrícula e saiam da página.

Assim, neste subcapítulo de descrição da Etapa II, serão novamente expostos os resultados da página de referência e, depois, os resultados obtidos

com os estudos da campanha e palavras utilizadas no Google AdWord. Posteriormente, foram avaliadas as métricas relativas a páginas de entradas, agrupamentos e conversa, até o momento em que o usuário deixa a página. Também na Etapa II, analisa-se a página formulário de matrícula e alguns aspectos do marketing relacionado ao seu preenchimento “com sucesso” e a análises de funil. Finalmente, tem-se a análise fatorial com foco de redução de estágios.

#### 4.4.1 Exame da campanha publicitária do Google Adwords

Durante o exame dos dados da campanha publicitária Google Adwords, observou-se (Figura 32) que menos de 9% entram no curso ESL utilizando uma máquina de busca.

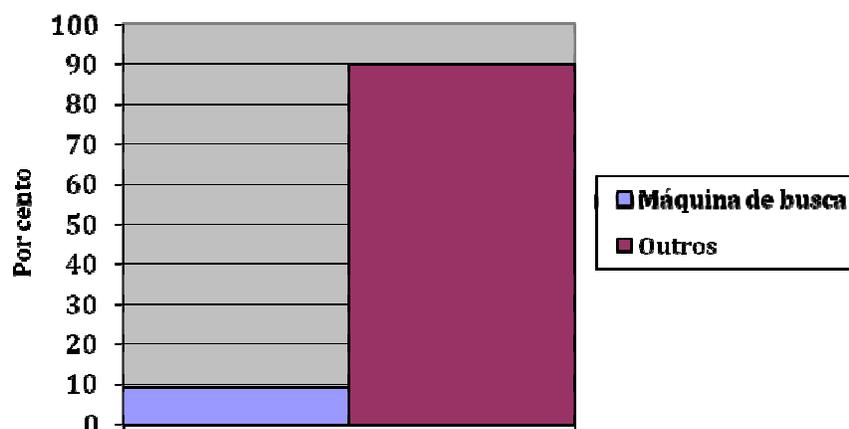


Gráfico 9 Uso da máquina de busca – Google  
Fonte: Track4Mine

Tem-se que a campanha no Google Adwords, com as palavras definidas pela organização (Quadro 6), não surte efeito. Neste quadro, quando são

utilizados "---", significa que a palavra-chave nunca foi encontrada (usada por visitantes) em todo o tempo de medição. Enfim, no período da medição ou de realização desta pesquisa, não foram utilizados as palavras-chave definidas pela organização. A propaganda, portanto, não tem o adequado retorno.

Quadro 5 Palavras-chave Adwords usadas para acesso ao curso ESL.

Palavras-chave	Número de variáveis	Número de uso
<b>computação ufla</b>	<b>1110</b>	<b>0</b>
<b>curso de engenharia de software</b>	---	---
<b>engenharia de software</b>	---	---
<b>engenharia de software experimental</b>	---	---
<b>engenharia de software requisitos</b>	---	---
<b>engenharia de software uml</b>	---	---
<b>graduação engenharia de software</b>	<b>768</b>	<b>0</b>
<b>mestrado ciência da computação</b>	---	---
<b>mestrado computação</b>	<b>659</b>	<b>0</b>
<b>mestrado computação</b>	<b>1265</b>	---
<b>mestrado sistemas de informação</b>	---	---
<b>pós-graduação computação</b>	<b>1059</b>	<b>0</b>
<b>pós-graduação informática</b>	---	---
<b>pós-graduação informática</b>	---	---
<b>produção de software</b>	---	---

Fonte: Track4Mine

Enfim, as palavras-chaves utilizadas e definidas pela organização não são aquelas utilizadas efetivamente pelas pessoas que realizam as buscas utilizando o Google.

As palavras-chave que são listadas no Quadro 6 são o resultado da pesquisa orgânica (geradas pelo próprio mecanismo de busca do Google e não a partir de registro da organização). Com estas palavras, sim, o cliente encontrará a página da UFLA. Estas palavras-chave mostram em que palavras os usuários têm um grande peso, quando eles procuram uma página da internet.

Quadro 6 Palavras-chaves utilizadas na máquina de busca – modo orgânico

<b>Palavra-chave</b>	<b>Número de uso</b>
Nenhuma palavra-chave	2595
ESL UFLA	7
Software livre	6
UFLA engenharia de software	5
UFLA esl	4
Engenharia de software	4
Curso engenharia de software	4
Engenharia de software UFLA	3
Outros	99

Fonte: Track4Mine

#### 4.4.2 Exame da página de referência

A origem de entrada do usuário na página do website ESL é variável, ou seja, ele pode vir de qualquer página, denominada página de referência. Assim, os visitantes podem entrar, direta ou indiretamente, no ambiente do website ESL. Os links de entradas podem ser, por exemplo, uma máquina de busca, e-mail, websites de redes sociais ou quaisquer outras páginas.

Na avaliação deste trabalho, foram cinco os principais elementos de referência: (1) acesso direto, (2) via domínio da UFLA, (3) máquina de busca, (4) rede social e (5) e-mail, conforme ilustrado na Figura 32.

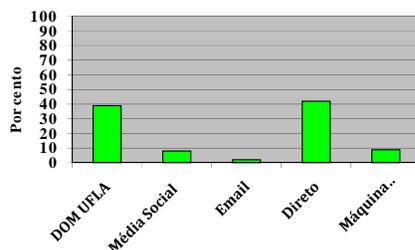


Gráfico 10 Gráfico das referências de entrada no website ESL  
Fonte: Track4Mine

Tem-se que, do grupo de máquina de busca, os usuários são provenientes da página principal Google (<http://www.google.com.br>) ou da página de busca especial, oferecida também pela Google, ou seja, o URL <http://www.images.google.com.br>. O grupo referencial menor é o originário de e-mail, particularmente das URLs Google: <http://mail.google.com> e Yahoo: <http://br.mg2.mail.yahoo.com>. Finalmente, observou-se que a grande maioria dos usuários é originária dos sites da própria universidade, ou domínio UFLA ([www.ufla.br](http://www.ufla.br)) e ou acessa diretamente o endereço ESL ([www.nte.esl.ufla.br](http://www.nte.esl.ufla.br)).

#### 4.4.3 Exame taxa de retorno (bounce rate)

A medição da taxa de retorno (Tabela 5) expressa a porcentagem de visitantes que deixaram a página após o primeiro clique. Esta análise é utilizada pela organização para a avaliação do valor de informação do website e permite

detectar se a propaganda inserida combina com o site e o valor que o usuário dá ao seu conteúdo. Cem por cento indica que todos os usuários deixaram a página após o primeiro clique.

Tabela 5 Taxa de retorno - Etapa II

ETAPA	ENTROU	UM (1) CLIQUE	TAXA DE RETORNO (%)
Home	2299	1436	62,54
O curso	0084	0065	77,38
Metodologia	0038	0031	81,16
Objetivos	0020	0015	75,00
Disciplinas	0176	0032	75,00
Professores	0024	0021	87,50
Promoção Matrícula	0117	0065	55,55
<b>Contato</b>	0020	0015	75,00
Credenciamento	0007	0004	57,15
<b>Formulário</b>	0031	0017	54,83
<b>Confirmação</b>	0005	0002	40,00

Fonte: Track4Mine

Conforme já observado anteriormente, em linhas gerais pode-se tomar por base o seguinte critério de avaliação:

- a) **21%-50%:** resultados obtidos pela ampla maioria das organizações, considerados normais. É possível melhorar o *bounce rate* adotando técnicas de otimização, além de focar em seu nicho específico;
- b) **51%-100%:** estes resultados são extremamente desfavoráveis. Sua taxa de retorno está muito elevada; os visitantes estão repudiando o que está sendo apresentado.
- c) Verifica-se que as páginas formulário de matrícula e confirmação são as melhores colocadas, mas deve-se considerar o baixo número de acessos a essas duas páginas.

#### 4.4.4 Agrupamento das variáveis do usuário

Nesta métrica, procurou-se classificar e agrupar alguns elementos de uso da web pelo visitante, tais como soma dos cliques, tempo gasto e resolução do navegador. Dessa maneira, foram identificadas semelhanças de comportamento de visitantes, o que permite sua classificação em segmentos. Os usuários do website têm certo número de visitas, ou melhor, de frequência de cliques na navegação pelo website (Figura 33). Dessa forma, cada visitante executa, no mínimo, um ou, então, mais cliques por visita. A frequência dos cliques permite calcular o percentual da taxa de atividade pela divisão do número de visitantes de um clique pelo total de visitantes multiplicado por 100. Por outro lado, verifica-se, nitidamente, quando os visitantes, de clique a clique, deixam o site. Nos casos em que, após determinada mudança de informações, conteúdo ou *layout* do website, houver queda na frequência dos cliques, fica claro que o novo conteúdo é de menor atratividade e valor para o visitante.

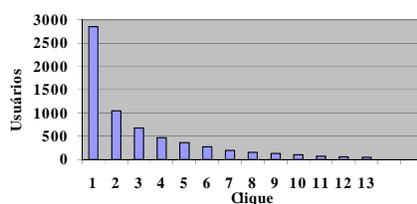


Gráfico 11 Agrupamento de cliques – Etapa II  
Fonte: Track4Mine

Outra forma de segmentação é a classificação dos usuários do site em grupos, demonstrando o tempo total da visita por colunas escuras e o tempo médio do site em colunas claras (Figura 34).

No caso de um tempo médio ou total da visita menor do que 20 segundos, pode se afirmar que o visitante pouco se dedicou às informações escritas ou à leitura do conteúdo do site. Se, após uma mudança do conteúdo do site, os tempos apurados mudam, pode-se aferir se houve melhora no tempo dedicado e, então, houve maior interesse e dedicação, ou vice-versa.

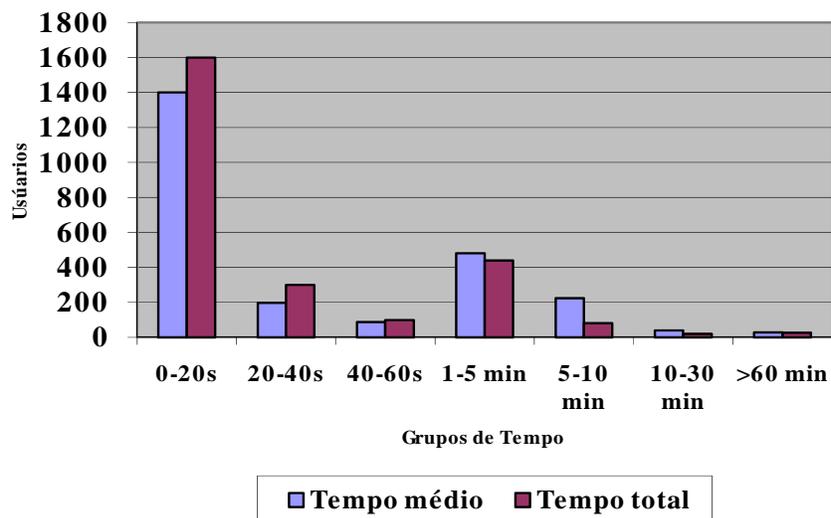


Gráfico 12 Agrupamento por tempo – Etapa II  
Fonte: Track4Mine

Os dados também permitem levantar (1) o sistema operacional, (2) o tipo de navegador, (3) a resolução e (4) a segmentação demográfica.

Os sistemas operacionais preferidos pelos visitantes são: Windows Vista e Windows XP (Figura 35). Há também usuários que utilizam os sistemas Winserver, Win2000 e MacOS.

Com relação aos navegadores, os mais utilizados são Mozilla Firefox e Internet Explorer (Figura 36).

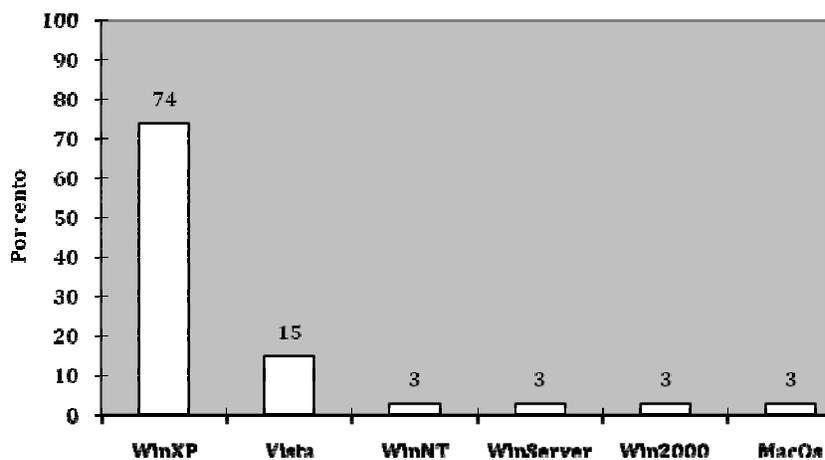


Gráfico 13 Sistemas operacionais utilizados para acessar a página ESL  
Fonte: Track4Mine

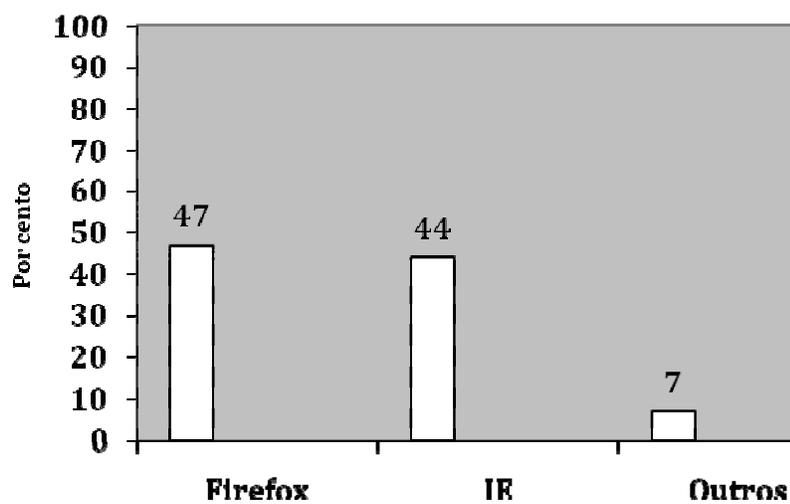


Gráfico 14 Navegadores utilizados para acessar a página ESL  
Fonte: Track4Mine

A resolução mais utilizada (Figura 37) é 1220 x 1024 pixels e 800 x 768 pixels. Esta análise é útil, principalmente para testes de atualização, uma vez que se pode redesenhar o website visando otimização para as principais preferências.

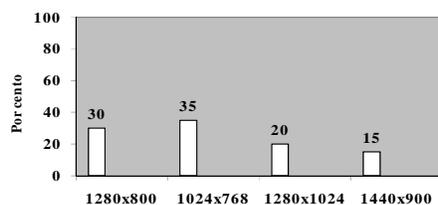


Gráfico 15 Resoluções mais utilizadas para acessar a página ESL  
Fonte: Track4Mine

O estudo da segmentação demográfica apresentado na Figura 38 ilustra que a maioria dos visitantes da página é do Brasil. Grande parte encontra-se em São Paulo, Distrito Federal, Paraíba, Rio de Janeiro e Minas Gerais.

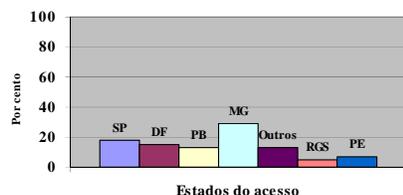


Gráfico 16 Estado de origem dos visitantes da página ESL.  
Fonte: Track4Mine

#### 4.4.5 Exame da atividade

A taxa de retorno de medidas (*bounce rate*), principalmente se os visitantes estiverem interessados, não significa necessariamente que se interessam pelo conteúdo de uma página, poderia ser por uma imagem ou outra coisa. Com o exame de atividade, o pesquisador pode ver, mais especificamente, por qual página as pessoas têm mais interesse.

Por exemplo, na Tabela 6 são mostrados os primeiros vinte cliques de todos os casos. Considerando o primeiro clique, tem-se que a concentração de cliques encontra-se nas páginas P1 (home), P7 (propaganda da matrícula) e P5 (disciplina). Do segundo até o sétimo clique, a maioria das pessoas está visitando a página P2 (curso) e a página P5 (disciplina). E, nos próximos cliques, elas, principalmente, querem mais informações sobre a página P5 (disciplina) e a página P6 (professor). Menos interesse eles tiveram na página P9 (credenciamento). É interessante observar o que a maioria das pessoas faz depois do terceiro clique, uma visita ao formulário, mas não fazem a matrícula.

Tabela 6 Exame da atividade - Etapa II

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
<b>Clique 1</b>	2299	084	038	020	176	024	117	020	007	031	005
<b>Clique 2</b>	0114	273	048	016	187	031	241	072	020	034	007
<b>Clique 3</b>	078	074	104	033	126	027	067	042	014	105	009
<b>Clique 4</b>	053	042	039	065	087	026	053	028	006	040	028
<b>Clique 5</b>	050	023	019	029	126	032	025	009	004	028	020
<b>Clique 6</b>	035	024	016	016	088	044	018	008	006	012	008
<b>Clique 7</b>	017	022	012	007	055	034	019	014	004	010	004
<b>Clique 8</b>	016	009	004	014	054	025	012	008	005	009	001
<b>Clique 9</b>	026	007	013	009	039	016	007	006	000	006	003
<b>Clique 10</b>	011	007	003	007	034	015	012	003	003	004	002
<b>Clique 11</b>	000	007	005	006	033	006	002	004	002	007	000
<b>Clique 12</b>	006	003	007	001	025	012	008	002	002	001	003
<b>Clique 13</b>	007	004	002	004	018	010	004	003	001	006	001
<b>Clique 14</b>	004	003	003	002	014	006	002	003	000	002	001
<b>Clique 15</b>	002	000	003	002	014	006	002	000	000	002	001
<b>Clique 16</b>	002	005	001	001	011	004	001	002	001	001	001
<b>Clique 17</b>	000	001	000	000	009	006	000	001	000	001	001
<b>Clique 18</b>	002	000	001	000	006	006	000	002	001	001	000
<b>Clique 19</b>	002	002	000	001	005	001	001	001	001	001	001
<b>Clique 20</b>	003	000	002	001	002	003	000	002	000	000	001

Fonte: Track4Mine

Na Tabela 7 constam as atividades efetivas no segundo período. Esta Tabela refere-se aos resultados da Tabela 6. Maior impacto tem a página P1 (home) e, depois, a página P5 (disciplina) e a página P7 (propaganda da matrícula). Está claro que os visitantes têm muito interesse na página P5 (disciplina) e pouco na página P9 (credenciamento).

Tabela 7 Taxa de retorno na Etapa II

Fase	Cliques	Exame da atividade
Home	<b>2738</b>	<b>41,12</b>
Curso	<b>570</b>	<b>8,56</b>
Metodologia	<b>360</b>	<b>5,41</b>
Objeto	<b>251</b>	<b>3,77</b>
Disciplina	<b>1110</b>	<b>16,67</b>
Professor	<b>353</b>	<b>5,30</b>
Propaganda da matrícula	<b>572</b>	<b>8,59</b>
Contato	<b>203</b>	<b>3,1</b>

"Continua"...

“Tabela 7, continuação”

Fase	Cliques	Exame da atividade
Credenciamento	45	1,42
Formulário de matrícula	271	4,07
Confirmação	133	2,00
Impressos	6656	100

Fonte: Track4Mine

#### 4.4.6 Frequência da próxima fase

A navegação pelo site também pode ser analisada por meio de regras de associação. Cada usuário tem um caminho, um clique em especial, que é transferido para uma matriz. No Quadro 7 pode-se observar que o usuário 4 entrou pela página de número quatro, que é a página de propaganda de matrícula (P7). Em seguida, passou pela sequência: curso (P2)=>metodologia (P3)=>objetivo (P4)=>disciplina (P5)=>professor (P6).

Quadro 7 Exemplo da intercomunicação da população nesta pesquisa

USUÁRIO	PN1	PN2	PN3	PN4	PN5	PN6
Usuário 4	P7	P2	P3	P4	P5	P6
Usuário 5	P1	P2	P1	P2	P7	P2

É possível, a partir desta navegação, tentar prever qual página o usuário irá acessar no próximo clique. Enfim, cada atividade do usuário é guardada em uma matriz, como na Tabela 4. Por exemplo: usuário 5 vem de fora (start) e entra na página inicial home (P1) (coluna), adicionando 1 para a coluna P1 em linha 'start'. Então, ele vai de home (P1) para curso (P3), adicionando 1 para a coluna curso P3 na linha home e assim por diante, até que o caminho do clicar seja acabado. Todos os números das páginas estão listados no capítulo 4.1, arquitetura do site ESL.

Quando todos os caminhos do usuário são salvos na matriz, cada célula é dividida pela sua soma total de linha, resultando em percentual ou

possibilidade de o usuário clicar em um link (coluna), quando estiver em uma página específica (linha).

Pela matriz representada na Tabela 7, fica claro que a maioria das pessoas entra nas páginas P1 (home), P5 (disciplina) e P7 (propaganda da matrícula). Os números das linhas mostram onde o usuário irá, na próxima vez. Assim, pode construir o caminho principal do percurso de usuário.

Os caminhos principais, ou seja, aqueles que são mais utilizados pelos visitantes do website ESL são:

- a) P1 (home) => P2 (curso) => P3 (metodologia) => P5 (disciplina) => P6 (professor) => fim;
- b) P1 (home) => P2 (curso) => P7 (propaganda da matrícula) => P8 (contato) => P10 (formulário) => P11 (confirmação) => fim;
- c) P1 (home) => P7 (propaganda da matrícula) => P8 (contato) => P10 (formulário) => fim;
- d) P1 (home) => P7 (propaganda da matrícula) => P8 (contato) => P10 (formulário) => P11 (confirmação) => fim.

Tabela 8 Matriz de frequência na próxima fase na etapa II

PAG.	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	Fim
<b>Iniciar</b>	81,5	2,9	1,3	0,7	6,2	0,8	4,1	0,7	0,2	1,1	0,1	
<b>P1</b>	1,0	12,5	1,9	1,2	7,2	1,3	10,4	1,4	1,1	0,5	0,0	60,9
<b>P2</b>	10,0	1,7	20,	2,4	11,8	1,1	21,2	1,1	0,1	0,4	0,2	29,5
<b>P3</b>	8,0	8,3	1,9	32,4	13,8	2,5	4,1	0,3	0	0	0,6	27,6
<b>P4</b>	6,6	7,1	8,0	3,3	43,6	4,7	3,3	0,9	0,4	0	0	21,8
<b>P5</b>	5,2	4,8	2,7	2,4	36,3	12,8	2,1	8,1	0,2	0	0,9	24,1
<b>P6</b>	0,9	7,8	1,6	6,2	18,0	23,0	8,5	2,9	0,9	0,3	0,3	28,9
<b>P7</b>	6,7	4,3	1,7	0,4	2,7	0,9	1,4	9,6	0,6	44,4	0,6	26,0
<b>P8</b>	15,0	6,3	3,4	0,5	5,2	4,6	9,8	1,7	11,5	4,6	0,5	36,4
<b>P9</b>	22,9	8,3	2,0	4,1	10,	8,3	8,3	20,8	0	8,3	0	19
<b>P10</b>	15,4	2,4	0,7	0,7	1,0	0,7	8,0	4,9	0	2,1	23,8	40,0
<b>P11</b>	13,9	7,5	8,6	3,2	11,8	1,0	5,3	3,2	1,0	3,2	5,3	35,4

Fonte: Track4Mine

#### 4.4.7 Análises de funil em cima dos grupos

Neste subcapítulo, consta a análise de funil considerando diferentes grupos de visitantes e respectivos registros. Esses grupos foram selecionados a partir de dados de tempo médio, do tempo total (Tabela 8), do número de cliques que um usuário fez (Tabela 9) e dos referencias URL que os usuários entraram no website ESL.

O funil foi construído a partir da entrada ate o destino, a página de formulário de matrícula também foi monitorada, visando conhecer se o usuário está confirmando ou não confirmando o formulário.

O funil de grupo mostra que os visitantes que entram no website ESL são dos grupos membro UFLA e Direto. Os usuários que entram diretamente, ou grupo direto, tem acesso menor à página formulário de matrícula que os membros que entram a partir do endereço UFLA, basicamente 75% menos. Por outro lado, os grupos membros de origem e-mail e mídia social nunca confirmaram a página de formulário de matrícula.

Tabela 9 Análises de funil de grupos - elementos de referência fase - etapa II

<b>Grupo de referência</b>	<b>Entram</b>	<b>Formulário</b>	<b>Atingir</b>
<b>UFLA</b>	1096	132	21
<b>MÉDIA SOCIAL</b>	154	7	0
<b>MÁ. DE BUSCA</b>	286	29	4
<b>E-MAIL</b>	39	3	0
<b>DIRETO</b>	1244	54	7
<b>TOTAL</b>	2808	225	32

Fonte: Track4Mine

O funil do grupo do tempo total (Tabela 10) mostra que a maioria dos usuários gasta menos de 20 segundos no tempo total. Um número reduzido de usuários com esse tempo entrou na página formulário de matrícula. Se os

usuários gastam, por clique, entre 20 segundos a 5 minutos, eles entram mais no formulário, mas não confirmam este formulário.

Tabela 10 Análises de funil de grupos - tempo total por clique - Etapa II

<b>Tempo total por clique</b>	<b>Entram</b>	<b>Formulário</b>	<b>Atingir</b>
<b>0-20 s</b>	1487	16	2
<b>20-40 s</b>	184	12	2
<b>40-60 s</b>	132	9	0
<b>1-5 min</b>	496	56	3
<b>5- 10 min</b>	185	36	5
<b>10-30 min</b>	193	57	12
<b>30-60 min</b>	64	21	6
<b>&lt;60 min</b>	78	19	1
<b>TOTAL</b>	<b>2808</b>	<b>225</b>	<b>32</b>

Fonte: Track4Mine

Os usuários que gastam entre 5 a 60 minutos por clique entram muito no formulário e atingiram mais a confirmação do mesmo.

Assim, o funil do grupo tempo total por clique (Tabela 10) mostra que a maioria dos usuários gasta menos de 20 segundos por clique. Foram poucos os usuários deste grupo que entraram na página de formulário e menos atingiram a confirmação deste formulário. Se os usuários gastam, por clique, entre 20 a 40 segundos, eles entram mais no formulário, mas também não atingiram confirmam o mesmo.

Na análise de funil de grupos número de cliques (Tabela 11), os usuários que gastam entre 5 a 10 minutos por clique entram muito no formulário de matrícula e um grande número confirma seu envio. Os usuários que gastam mais de 10 minutos por clique nunca o confirmam.

Tabela 11 Análises de funil de grupos - tempo total por clique - Etapa II

<b>Tempo médio por clique</b>	<b>Entram</b>	<b>Formulário</b>	<b>Atingir</b>
0-20 s	1613	44	4
20-40 s	324	33	3
40-60 s	196	30	4
1-5 min	438	82	18
5- 10 min	100	21	3
10-30 min	79	10	0
30-60 min	31	4	0
<60 min	38	2	0
<b>TOTAL</b>	<b>2808</b>	<b>225</b>	<b>32</b>

Fonte: Track4Mine

Nas análises de funil de grupos, o número de cliques deixa claro que muitos usuários gastam menos de 4 cliques para chegar à página formulário de matrícula, mas poucos entram nesta página e um menor número preenche e encaminha o formulário. Um grupo de usuários gastou 5 cliques para confirmar e outro grupo gastou mais de 9 cliques. Isso pode significar que alguns usuários entendem ter as informações adequadas e outros necessitam mais informação.

Tabela 12 Análises de funil de grupos - número de cliques - Etapa II

Número clique	Número	Formulário	Atingir
1	1805	0	0
2	330	15	1
3	200	48	0
4	116	24	1
5	86	32	7
6	71	18	1
7	46	15	2
8	32	11	1
9	26	8	3
10	25	11	4
11	14	7	0
12	8	4	2
13	18	10	3
14	7	4	1
15	3	2	0
16	8	4	2
17	3	2	1
18	5	2	1
19	0	0	0
20	3	1	1
21	0	0	0
22	3	2	0
23	5	4	0
24	3	1	1
25	0	0	0
<b>TOTAL</b>	2808	225	32

Fonte: Track4Mine

#### 4.5 Discussão e propostas de estratégias – marketing e web mining

Os elementos de análise no contexto do marketing (propaganda) envolvendo o curso a distância ESL foram, basicamente, aqueles relacionados a propaganda *offline*, como folders, jornais, revistas, rádio e TV, e propaganda *online* considerando-se Google Addwords e o website ESL. A estratégia de ação da organização é a de marketing de massa, em todo o país.

Nas etapas I e II iniciou-se com o exame dos endereços de referência. Estes endereços são, principalmente, os de domínio da UFLA e direto. Isso pode

indicar se a propaganda *offline* está dentro de uma faixa adequada de eficiência, logicamente no sentido de gerar, no subjetivo do provável cliente, o interesse em visitar a página do curso.

Os outros grandes grupos são os de mídia social, máquinas de busca e e-mail. Sobre eles não foi possível ter uma visão conclusiva. Os grupos de máquina de busca deixam claro que eles buscam algo, mas os elementos de busca, as palavras, por exemplo, não têm sido eficientes. Finalmente, com relação à mídia social, o acesso é reduzido.

Especificamente no caso das mídias sociais, a visão que se tem é de que o investimento nessas mídias pode ocorrer, desde que as mesmas sejam relacionadas ao ensino ou a cursos. Por exemplo, se as pessoas estão discutindo sobre cursos ou temas parecidos com o do curso e ensino-aprendizagem. Já no e-mail, eles podem receber o *hyperlink* de um amigo e, por esse motivo, entrarem na página, como também por muitas outras razões. Isso sem analisar a intercomunicação entre esses, o que é muito complexo.

Nas etapas I e II, considerando a propaganda Google Adwords, tem-se que as palavras-chave que o marketing da organização havia escolhido para a propaganda não foram utilizadas. As palavras que foram encontradas são resultados da busca orgânica e não de palavras de origem organizacional. Mesmo assim, observou-se que poucas pessoas utilizam máquina de busca, ou seja, são originárias de máquinas de busca. No entanto, isso também pode querer dizer que muitos podem estar buscando o curso, mas, em função das palavras base e do próprio conjunto de palavras de uso geral, estes não estão encontrando. Uma possibilidade poderia ser buscar outras palavras-chave em combinação com as que efetivamente promoveram sucesso de busca.

Nas etapas I e II, a maioria das pessoas entra na página home e página de propaganda. O home não é sempre o melhor lugar em que os usuários poderiam entrar em uma página de negócios na web. Este estudo de caso pode

ser uma exceção. Em função dos poucos links, foi uma ideia usar a home ou ir direto a propaganda. As outras páginas não são muito representativas porque têm poucas entradas.

No agrupamento das variáveis nas etapas I e II, tem-se que são poucas as pessoas que permanecem por mais que oito cliques na página. Isso pode significar que eles não precisam ou não encontram muitas informações ou, talvez, que eles tomam a decisão antes de ver a informação *online*. A análise de tempo mostra que são poucas as pessoas que gastam mais de dez minutos em cada página. A maioria gasta menos de cinco minutos.

Os outros agrupamentos, como sistema operacional, navegador e resolução da tela, são nas etapas I e II, quase idênticos. Assim, estes valores poderiam ser utilizados quando uma nova página ou site for construído. Neste site não se tem muita utilização para estes dados, devido ao fato de todas as páginas utilizarem a mesma tecnologia.

No agrupamento por estado, os resultados indicam que é de estados próximos, como Minas Gerais, Rio de Janeiro e São Paulo, que vem a grande maioria dos visitantes. A equipe de marketing da organização tem que analisar se as revistas ou outras propagandas são mais centradas nestas regiões ou existem outras razões pelas quais estes estados acessam mais a página.

No exame das etapas I e II a maioria quer saber, depois do segundo clique, algo sobre o curso, a disciplina e os professores. Nesta página, em que muitas pessoas ficam, é um bom lugar para se incluir novas propagandas.

No estudo de clique da próxima fase, tem-se o percurso do caminho básico até o usuário encontrar um destino. Na Etapa I não se pode afirmar nada porque sem destino o caminho principal não reflete se o usuário tem interesse ou não. Mas, os caminhos principais podem colocar algo para que o consumidor encontre mais facilmente o caminho para o destino, neste caso, informações

sobre o curso. Mas, isso vai ser difícil porque a decisão para se matricular no curso é, provavelmente, uma decisão racional e menos uma habitual.

Na Etapa I não pode ser avaliado o destino da página formulário de matrícula porque não estava disponível, considerando a navegação e a avaliação de cliques.

Na Etapa II tem-se o monitoramento porque as análises do funil a partir de grupos permitem esse tipo de acompanhamento. Na primeira, a análise permitiu perceber que pessoas que entraram originárias de mídias sociais ou de e-mail nunca fizeram pelo menos uma inscrição. Na máquina de busca são poucos. Nas análises do funil, considerando tempo total, pode-se notar que usuários que preenchem o formulário gastam mais de cinco minutos no cômputo total. São muitos os visitantes que gastam menos de cinco minutos. Desses, poucos preenchem a página formulário de matrícula. Nas análises de funil considerando tempos médios, visitantes gastam entre quarenta segundos e dez minutos para alcançar e preencher a página formulário de matrícula”. Finalmente, nas análises de funil de cliques, tem-se um grupo que gasta cinco cliques para alcançar a página formulário de matrícula, tendo as pessoas que gastam mais que nove cliques demonstrado mais interesse em ler e preencher o formulário.

## 5 CONSIDERAÇÕES FINAIS

Neste trabalho, procurou-se investigar a tecnologia de mineração de uso da web (*web usage mining* ou *web mining* de uso), definida como o processo que explora e registra o comportamento do navegar de visitantes em páginas de um website, com ações, práticas e atividades de marketing e relacionamento com clientes desenvolvidas por uma organização da área de prestação de serviços educacionais.

Com a pesquisa, fundamentada em análise documental e métricas *web mining* aplicadas aos dados coletados, pode-se concluir que a organização não se encontra adequadamente preparada para o processo de *web mining*, considerando, principalmente, que seu plano de propaganda é quase que totalmente baseado em artefatos de difusão de massa no modo *offline*, ou seja, folders, revistas, TV, jornais e rádios. O uso de componentes de marketing *online* é limitado ao website da organização, do curso e a acordos comerciais envolvendo o Google Adwords e máquinas de busca, ou seja, onde os visitantes realizam busca no Google, a partir de termos registrados pela organização ou palavras-chave do usuário.

No entanto, mesmo com este contexto limitado, foi possível a aplicação de técnicas de *web mining* de uso ao website do curso de pós-graduação lato sensu em Engenharia de Software Livre (ESL) e a consequente discussão dos resultados envolvendo métricas e observações relacionadas a páginas de referência de acesso do visitante (páginas de origem do visitante), exame de taxa de retorno, agrupamento de variáveis do usuário (número de cliques, tempos de acesso, tipos de sistemas operacionais e navegadores utilizados, estudo demográfico ou de origem dos usuários, exame de atividade, frequência da próxima fase, além de uma abordagem de métricas estatísticas preliminares).

Estas análises, quando relacionadas ao procedimento real de propaganda e marketing da organização, contribuem para o entendimento teórico e prático das aplicações do *web mining*. Observou-se que, para que ocorra esta aplicação, os planos de marketing (propaganda) devem, antes, buscar o conhecimento e a interação com os componentes de arquitetura do website e com os profissionais a ele relacionados. Não há como partir para o web marketing sem o adequado conhecimento, planejamento, visão e arquitetura adequados ao website.

Nesse sentido, a pesquisa se deparou com dois problemas básicos: (1) na primeira fase, o formulário de inscrição não foi monitorado e (2) nas etapas I e II, a única propaganda *online* real, a do Google AddWords, foi um fracasso. São resultados ruins e que chamam a atenção para aspectos realmente necessários e efetivos de planejamento e profissionalismo, quando se trabalha com a web.

Para as organizações que queiram realizar *web mining*, aqui se encontra uma experiência não muito positiva, ou seja, do que não deve ser feito. Um próximo plano de propaganda e marketing deveria, por exemplo, se concentrar em ajuste de *design* ou projeto de páginas, inter-relações entre páginas, visão de funil – direcionar o visitante para o preenchimento de dados e possibilidades de contatos. Além disso, deve-se relacionar propaganda *offline* com *online*, potencializando acessos diretos à página de curso, considerando o estudo adequado de palavras de busca que ofereçam eficiência e eficácia no que se refere a encaminhar o consumidor ou mostrar a ele, em primeiro plano, a página do curso.

Por outro lado, as diferentes métricas e estudos realizados, apresentam uma contribuição empírica relacionada aos seguintes fatores: Que estudos podem ser efetivamente implementados considerando os dados coletados? É, sem dúvida, possível ter-se toda uma estratégia de marketing e relacionamento, mesmo o relacionamento pós-venda ou pós-contato, totalmente fundamentado na web. Os dados e as métricas permitem a descoberta de fontes ocultas do

comportamento do navegar web do visitante que podem, inclusive, mudar todo um complexo plano de propaganda e marketing. Além disso, as páginas do website necessitam permanecer dinamicamente acompanhadas pela equipe responsável, visando obter conhecimento para a adequada tomada, se possível, dinâmica de decisão.

A partir dos dados coletados podem ser construídas matrizes complexas de comportamento, de sequência, de agrupamento e de atividades, viabilizando a descoberta do conhecimento sobre o visitante. Enfim, as informações geradas, com a utilização da mineração de uso da web, com certeza, podem compor um ciclo de dados, informação e conhecimento que potencializem pesquisa, desenvolvimento e tomadas de decisão em marketing, produtos e serviços nas organizações.

O fato é que, mesmo considerando as limitações do caso estudado, foi possível adquirir conhecimento sobre o visitante web e interpretar alguns de seus objetivos e expectativas com a tecnologia *web mining* e, mais ainda, podem-se vislumbrar aplicações, estudos e pesquisas futuras, envolvendo a mineração de dados centrada no cliente (*customer centric data mining*).

## REFERÊNCIAS

ASH, T. **Landing page optimization**: definitive guide to testing and tuning for conversions. Hoboken: J. Wiley, 2008.

BRAGA, L. P. V. **Introdução à mineração de dados**. 2. ed. Rio de Janeiro: E-papers, 2005.

BURBY, J.; ATCHISON, S. **Web analytics**: using data to make smart business decisions. Köln: Sybex, 2007.

COOLIDGE, F. L. et al. On the relationship between Karen Horney's tripartite neurotic type theory and personality disorder features. **Pergamon**, Oxford, v. 30, p. 1387-1400, 2000.

ELEY, B.; TILLEY, S. **Online marketign inside out**. Melbourne: SitePoint, 2009.

EMERICK, D.; ROUND, K. **Exploring web marketing & projekt management**. New Jersey: Prentice Hall, 2000.

ENGEL, J. F.; BLACKWELL, R. D.; MINIARD, P. W. **Comportamento do consumidor**. 8. ed. Rio de Janeiro: LTC, 2000.

ENGENHARIA DE SOFTWARE LIVRE. Disponível em: < <http://www.nte.ufla.br/esl/wp>>. Acesso em: 12 jun. 2009.

HEIDRICHS, J. H.; LIM J. S. Integrating web-based Data Mining tools with business models for knowledge management. **Decision Support System**, Toletto, v. 35, p. 103-112, 2003.

HOLANDA, L. M. C. et al. Data webhouse: a evolução do data warehouse para Web e suas contribuições para o aperfeiçoamento do relacionamento com clientes. In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA, 3., Campina Grande. **Anais...** Campina Grande: SEGET, 2006.

HUANG, X. Comparison of interest measures for web usage mining: an empirical study. **World Science**, McLean, v. 6, p.15-41, 2007.

IRENE, S. Y. et al. An e-customer behavior model with online analytical mining for Internet Marketing planning. **Decision Support Systems**, Amsterdam, v. 41, p. 189-204, 2005.

KAUSHIK, A. **Web Analítica uma hora por dia Sybex**. Rio de Janeiro: J. Wiley, 2007

KRÖBER-RIEL, W. et al. **Konsumentenverhalten**. 9th ed. Munich: Vahlen, 2008.

LEDFOURD, J.; TYLER, M. E. **Google analytics 2.0**. Somerset: J. Wiley, 2007.

LEE, J. et al. Visualization and analysis of clickstream data of online stores for understanding Web merchandising. **Data Mining Knowledge Discovery**, Hingham, v. 5, p. 59-84, 2001.

MACHADO, L. S.; BECKER, K. O. uso da mineração de dados na Web aplicado a um ambiente de ensino a distância. In: WORKSHOP DE TESES E DISSERTAÇÕES EM BANCO DE DADOS, 1., 2007, Porto Alegre. **Anais...** Porto Alegre: PUCRS, 2007.

MARTIN, N. **Habit: the 95% of behavior marketers ignore**. New Jersey: FT, 2008.

MELLO, S. C. B.; LEAO, A. L. M. Uma revisão crítica sobre o uso de diferentes conceitos de valor no marketing. In: ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 2., 2008, Curitiba. **Anais...** Curitiba, 2008. p. 37-53.

MENASCÉ, D. A.; ALMEIDA, V. A. F. **Scaling for E-Business**. New Jersey: Predice Hall, 2000.

MONTEIRO, P. R. R.; VEIGA, R. T. Subcultura tuning: a identidade estendida na personalização de automóveis. In: ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 2., 2008, Curitiba. **Anais...** Curitiba, 2008.

MONTEIRO, P. R. R.; VEGA, R. T. Uma avaliação empírica do modelo metateórico de motivação e personalidade. In: ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 30., 2006, Salvador. **Anais...** Salvador, 2006.

MORAN, M. **Do it wrong quickly**: how the Web changes the old marketing rules. Upper Saddle River: IBM, 2007.

NGAI, E. W. T.; XIU, L.; CHAU, D. C. K. Application of data mining techniques in customer relationship management. **Expert Systems with Applications**, Amsterdam, v. 36 p. 2592- 2602, 2009.

OBERLE, V. **Seminar im wintersemester 1999/2000**. Disponível em: <<http://www.ipd.uka.de/~oosem/ECVU9900/Ausarbeitungen/VincentOberle.pdf>>. Acesso em: jun. 2010.

OPENUFLA. Disponível em: <[www.openufla.br](http://www.openufla.br)>. Acesso em: 12 fev. 2010.

PABARSKAITE, Z.; RAUDYS, A. A process of knowledge discovery from web log data. **Journal of Intelligent Information System**, Netherland, v. 28, p. 79-104, 2007.

REZENDE, S. O (Org.). **Sistemas inteligentes**: fundamentos e aplicações. Barueri: Manole, 2003.

SCHIFFMAN, L. G.; KANUK, L. G. **Consumer behavior**. 7. ed. London: Prentice Hall, 2000.

SILVA, L. C.; CERCHIARO, I. B.; MASCETTI, I. Consumo como espiritualidade e consume do luxo: o que o consumidor tem a dizer. In: ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 30., 2006, Salvador. **Anais...**Salvador, 2006.

SOLOMON, R. M. **O comportamento do consumidor comprando, possuindo e sendo**. 5. ed. Porto Alegre: Bookman, 2002.

SONG, J.; ZAHEDI, F. M. A theoretical approach to web design in e-commerce: a belief reinforcement model. **Management Science**, Hanover, v. 51 p. 1219-1235, 2005.

TRACK4MINE. Disponível em: <<http://tm-licesa.dcc.ufla.br>>. Acesso em: jun. 2009.

TUG, E.; SAKIROGLU, M.; ARSLAN, A. Automatic discovery of the sequential accesses from web log data files via genetic algorithm. **Elsevier**, Amsterdam, v. 19, p.180-186, 2005.

WIND, Y. J.; MAHAJAN, V.; GUNTHER, R. E. **Convergence marketing:** strategies for reaching the new hybrid consumer. New Jersey, 2001.

ZAIANE, O. M. **From resource discovery to knowledge discovery on the internet.** 1999. Thesis (Doctor of Philosophy) - School of Computing Science Simon Fraser University, Burnaby, 1999.