

WEB-MINING: CONCEITOS E APLICAÇÕES

Diogo Pietro Scoz
diogoscoz@gmail.com

Resumo

A grande quantidade e variedade de dados disponíveis na web cria a oportunidade da aplicação de técnicas de mineração de dados na Internet, isso é conhecido como *web-mining*. Quando uma pessoa necessita encontrar alguma informação na web, normalmente ela irá recorrer a algum motor de busca, e um dos maiores problemas é encontrar os dados relevantes dentre tantos resultados da busca de modo a conseguir extrair conhecimento útil. Por isso, são necessárias técnicas para aumentar a eficiência das respostas, visto que as pessoas possuem padrões diferentes de preferências e atividades que realizam online. Tais padrões e dados sobre os usuários podem ser “minerados” com o objetivo de tornar a experiência do próprio usuário em algum site mais eficiente. O objetivo deste artigo é explanar os conceitos básicos necessários para compreensão do tema de *web-mining*.

Palavras-chave: Internet. *Web-mining*. Conhecimento.

1. Introdução

Desde a popularização da Internet, as empresas investem na web, pois nelas veem soluções para diversas dificuldades, por exemplo de comunicação, compartilhamento de dados ou integração de conteúdo. Desta forma a web é vista como um meio de expor as informações, produtos e serviços para os clientes ou funcionários. Com a extensão do uso de aplicações web para as mais diversas atividades, as organizações perceberam a quantidade de conhecimento contido no uso, maneira, quantidade, ou até local de acesso de suas páginas pelos usuários, assim surgiu o conceito de conhecimento na web como forma das empresas conhecerem melhor o mercado qual estão inseridas (SRIVASTAVA, 2010).

A evolução das tecnologias de comunicação sem fio, processadores e dos dispositivos eletrônicos digitais de uso pessoal, tornou possível fornecer aos usuários, de qualquer lugar, a qualquer hora e usando qualquer tipo de dispositivo o fácil acesso a serviços e aplicações. Essas capacidades levaram ao surgimento do conceito de ubiquidade na computação (ARAÚJO, 2003).

Com toda a disponibilidade dos dados oriundos das transações online, o web mining permite as empresas irem ao encontro dos clientes oferecendo a eles aquilo que eles realmente precisam, pois baseiam-se na localização, preferências e perfil pessoal obtido a partir do comportamento online do usuário. Além de dados pessoais, o web mining opera sobre diferentes outras fontes como bancos de dados de organizações, servidores web e servidores Proxy (SRIVASTAVA, 2010).

O objetivo deste trabalho é evidenciar os considerados maiores impactos da web mining para os desenvolvedores web. Benefícios e críticas serão abordados, bem como uma breve conceituação dos temas necessários para a compreensão do tema como um todo. O restante do artigo está organizado desta forma: em seguida uma breve contextualização do tema. A seguir um aprofundamento no funcionamento e classificações de *web mining* bem como suas principais vantagens e críticas. Por fim as considerações finais.

2. Conceitos

A grande informatização de processos e de dados que ocorreu nas organizações desde a popularização dos computadores resultou no fato de muitas empresas possuírem massivos estoques de dados, como por exemplo o Walmart que realiza e registra mais de um milhão de transações por hora em mais de 2000 filiais (THE ECONOMIST, 2010). Com tamanho volume, estratégias tradicionais de tratar estes dados se tornaram menos efetivas na tarefa de transformar dados “crus” em conhecimento através do reconhecimento de padrões (SRIVASTAVA, 2010).

3. Web-mining

Segundo Cordeiro (2003) *web mining* é um termo que surgiu do DM que é uma área de estudo voltada à descoberta de conhecimento sobre bases de dados convencionais utilizando técnicas de procura como Inteligência Artificial. O relacionamento entre DM e *web mining* se dá no sentido de que ambas visam a descoberta de padrões úteis em conjuntos de dados (CORDEIRO, 2003).

Devido a vasta quantidade de diferentes aplicações, a *web mining* foi dividida em três categorias por Zaiane (1999) de acordo com seus campos de atuação: *web content mining*, *web structure mining* e *web usage mining*, como pode ser observado na figura 1

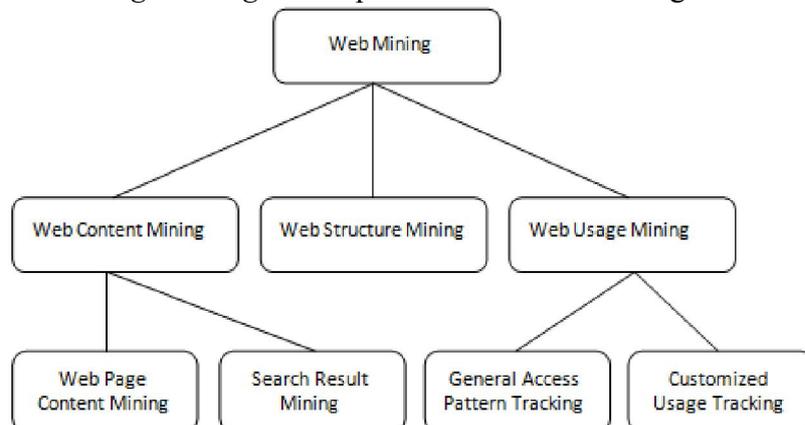


Figura 1 – Categorias de *web-mining*, adaptado de Zaiane (1999).

Categorizar este tema é uma forma de especializar o conhecimento individual necessário para o desenvolvimento eficiente de cada categoria (ZAIANE, 1999). A figura 1 ilustra como a categorização é realizada no *web mining*, por este mesmo motivo, o *web content mining* e o *web usage mining* também possuem divisões que representam especializações do conhecimento muito amplo contido nelas.

3.1 Web Content Mining

Considerando que os métodos tradicionais de busca se mostram ineficazes em face a realizar buscas na web, principalmente as empresas de mecanismos de busca, como a Google buscam desenvolver e utilizar técnicas mais eficientes para facilitar o processo de pesquisas na web. Estas técnicas são classificadas como *web content mining* (WCM) que segundo a definição de Zaiane (1999) representam o processo de extração de conhecimento do conteúdo de páginas.

A categoria de WCM se sub divide em *web page content mining* (WPCM) e *search result mining* (SRM). A diferença entre ambas é o fato de que WPCM, também conhecida como “extração de informação” busca por algum conteúdo específico relacionado com o tema informado como critério, enquanto o SRM busca por documentos ou páginas que se relacionem com os critérios solicitados (SRIVASTAVA, 2010).

3.2 Web Structure Mining

Segundo Zaiane (1999) *web structure mining* (WSM) pode ser conceituada como a atividade de inferir conhecimento da arquitetura e organização de páginas web e ligações entre referências de sites. Zaiane (1999) afirma ainda que a WSM utiliza de análises nos relacionamentos entre páginas feitos por *hiperlinks* para otimizar motores de busca. Essa otimização ocorre principalmente na redução da quantidade de resultados irrelevantes em buscas. Outra contribuição valiosa é o combate do problema de dificuldade de indexação da grande quantidade de dados encontrados na web, já que os sites passam a ser analisados como grafos.

Para Scime (2004) o uso da WSM pode contribuir estrategicamente nas atividades de marketing em sites de e-commerce. Considerando que a taxa de memorização de uma página pelos motores de busca se dá pela quantidade de redirecionamentos para esta página, e que isso indiretamente aumentará a quantidade de retornos da página de maneira cíclica.

3.3 Web Usage Mining

A *web usage mining* é definida por Zaiane (1999), quem criou o termo, como a atividade de descobrir automaticamente padrões de acessos de usuários em sites baseado em registros de páginas acessadas.

As grandes empresas que fornecem mecanismos de busca aos usuários, costumam manter em histórico as buscas realizadas por cada usuário, desta forma elas podem analisar estes dados, procurar por padrões e definir perfis pessoais que podem ser utilizados para as mais diferentes finalidades. Fornece uma visão realista das reais preferências do usuário, logo pode servir de auxílio no processo de geração de campanhas de marketing e direcionamento do público alvo (SCIME, 2004).

4. Aplicações

Por funcionar baseado em DM, o *web mining* possui uma vasta gama de possibilidades de aplicações (SRIVASTAVA, 2010; SANTOS, 2014). Em seus primórdios as ferramentas de *web mining* relatavam simplesmente algumas informações geradas nos sites, como o número de visitas, as páginas mais visitadas e a origem dos usuários. Entretanto, segundo Monteiro (2006) surgiram e continuam sendo criadas cada vez mais novas funcionalidades e ferramentas mais otimizadas para atuação em *web mining*. Monteiro (2006) acrescenta ainda que as ferramentas estão convergindo para uma busca por informações mais relevantes contextualmente e não apenas estatísticas.

Segundo Santos (2014), as ferramentas de *web mining* mais utilizadas em empresas norte americanas são: Clementine e SAS. Em análise realizada sobre as funcionalidades propostas por cada um desses aplicativos, pode-se conferir o grande número de possibilidades de aplicações como citado anteriormente. O Clementine oferece suporte a DM, *business analytics*, *business intelligence*, *predictive analytics*, além do próprio módulo de *web mining*. Enquanto o SAS tem uma abordagem mais direcionada ao marketing, segundo a própria desenvolvedora SAS (2002) o objetivo do sistema é responder perguntas do tipo:

- Quem são meus clientes?
- Quanto tempo um usuário costuma permanecer no site?
- Quando e de onde eles acessam?
- A frequência de compras dos usuários, e o que está relacionado com mais ou menos compras?
- O que é mais comprado?

Com as respostas ou pelo menos *insights* sobre as questões listadas acima, o desenvolvedor web possui uma base concreta de informações para se basear e tomar decisões sobre o formato do site. Como o SAS armazena grandes quantidade de *clickstreams*, que relatam a sequência de cliques e ações de cada usuário no site, o sistema permite a inferência de otimizações de acordo com as sequências mais realizadas, de modo a evitar passos desnecessários (SAS, 2014). Para maior detalhamento na comparação entre as duas ferramentas citadas, vide o trabalho de Santos (2014).

Considerações Finais

A criação de sites de qualidade para os padrões e requisitos modernos envolve importantes fatores destacados no decorrer deste trabalho, como sua estruturação, o formato de apresentação das informações em tela, caminhos de navegação otimizados e ser amigável aos motores de busca. Uma maneira de atingir eficientemente estes objetivos é através de *web mining*.

Web mining possibilita organizar a páginas para que sejam exibidos somente os recursos necessários, eliminando os dados que são irrelevantes para o usuário. A partir destas vantagens busca-se atender as verdadeiras necessidades dos usuários, as que são capturadas de maneira ubíqua através do seu comportamento no site, e assim criar estratégias de marketing mais adequadas e abrangentes.

A mineração de dados aplicada na web é vista como uma oportunidade de obter um diferencial de mercado, já que os sites podem ser adaptados para armazenar todos os comportamentos e padrões dos usuário, e estes dados são uma valiosa fonte de potencial conhecimento sobre as preferências dos consumidores. Naturalmente, conhecer melhor os clientes, permite às organizações tomarem decisões mais assertivas pelo fato de serem baseadas em informações concretas, e não apenas palpites ou guias genéricos de usabilidade.

Referências

CORDEIRO, J. 2003. **Extracção de Elementos Relevantes em Texto/Páginas da World Wide Web**. Tese de Mestrado em Inteligência Artificial e Computação. Faculdade de Ciências da Universidade do Porto. Disponível em: <<http://www.di.ubi.pt/~jpaulo/publications/MSc-JPC.pdf>> Acesso em: 02.11.2014.

MONTEIRO, M. 2006. **Web Usage Mining**. Disponível em: <http://pt.scribd.com/manuel_monteiro_18> Acesso em: 03.11.2014.

SANTOS, E. 2014. **Web Mining: Mineração de dados na Web utilizando CMS**. Monografia em Engenharia de Sistemas e Informática pela Universidade Jean Piaget de Cabo Verde. Disponível em: <<http://bdigital.cv.unipiaget.org:8080/jspui/bitstream/10964/543/1/Mem%C3%B3ria%20de%20Emanuel%20dos%20Santos%20ESI-SI.pdf>> Acesso em: 02.11.2014.

SAS. 2014. **SAS: Business Analytics and Business Intelligence Software**. Disponível em: <http://www.sas.com/en_us/home.html> Acesso em: 04.11.2014.

SRIVASTAVA, J. 2010. **Web Mining: Accomplishments & Future Directions**. University of Minnesota. Disponível em: <<http://www.ieee.org.ar/downloads/Srivastava-tut-pres.pdf>> Acesso em: 02.11.2014.

THE ECONOMIST. 2010. **Data, data everywhere**. Disponível em: <<http://www.economist.com/node/15557443>> Acesso em: 02.11.2014.

ZAIANE, O. 1999. **Resource and Knowledge Discovery from the Internet and Multimedia Repositories**. School of Computing Science, Simon Fraser University, Canada
Disponível em: http://www.collectionscanada.gc.ca/obj/s4/f2/dsk1/tape7/PQDD_0025/NQ51940.pdf > Acesso em: 02.11.2014.