

IF796 - Mineração na Web

Matheus Marinho De Moraes Leça
mmml2@cin.ufpe.br

2018

1 Introdução

De forma geral, a mineração na Web pode ser conceituada como a descoberta e análise inteligente de informações úteis da Web[1]. Essa ideia foi concebida a partir do desejo de filtrar, ordenar e recuperar informações que estão na World Wide Web[2].

As abordagens da Mineração na Web são: [3]

1. Mineração de conteúdo;
2. Mineração de estrutura;
3. Mineração de uso.

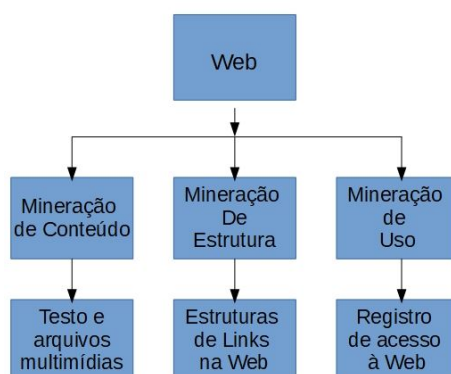


Figura 1: Diagrama esquematizando a atuação da Mineração da Web

A primeira cuida da informação contida dentro dos documentos da Web, a segunda da informação contida entre os documentos da Web e a última na informação contida na utilização ou interação com a Web[4].

Essas são as três categorias que divide a mineração da web.

2 Relevância

A Mineração na Web não é já tão recente, pois vem sendo estudada desde meados de 1996, porém sua relevância tem sido gradativamente aumentada ao passar dos anos. Podemos indicar duas principais razões para isso: [5]

- Aumento das transações comerciais na Web, que estimularam o desenvolvimento de técnicas de mineração de uso, pois assim entende-se melhor o perfil de seus usuários, melhorando no marketing e vendas;
- O desenvolvimento da Web semântica e da tecnologia dos agentes da informação, onde as técnicas de Mineração na Web são utilizadas. A Web semântica consegue entender as inteligências dos agentes e não apenas o seu comportamento. Assim, serviços da Web poderão ter uma interatividade melhor entre si por meio de uma linguagem comum. A Mineração na Web é crucial para isso, pois ajudará na busca de informações, personalização e talvez até como mecanismo de aprendizado.

Realmente o Web Mining, ou Mineração na Web, é uma poderosa ferramenta usada para descobrir padrões da Web, mas esta tecnologia também tem seus pontos negativos, além dos positivos, eles estão a seguir[5]:

2.1 Pontos positivos

- É uma ferramenta poderosa que desperta a atenção de muitas agências governamentais por oferecer uma grande quantidade de dados a fim de vigilância, no caso de uma agência de polícia, e a fim de estatísticas, a fim de certos órgãos governamentais.
- Ela também é muito utilizada no e-commerce para se personalizar as propagandas e assim despertar melhor o interesse do consumidor tendo como base seus próprios dados.
- Empresas também usam como um meio de estreitar sua relação com o consumidor e agilizar o processo de ajuste em seus produtos para suprir a demanda.

2.2 Pontos negativos

- A ferramenta em si não tem muitos problemas, mas sim seu mal uso. Denúncias de invasões de privacidade são recorrentes nesse meio justamente quando a extração dos dados não foram feitos com consentimento.
- Outro problema é quando as empresas alegam usar os dados para um determinado fim e usam em outro totalmente diferente, violando o consentimento do consumidor.
- Alguns algoritmos de mineração podem coletar esses dados para fins controversos, como para categorizar em função de orientação sexual, gênero, cor, religião e etc. Essas práticas podem estar ferindo alguma legislação de direitos humanos.
- E outro polêmico ponto é que as empresas que coletam esses dados são as que tem a legal posse delas, mas se caso algum ataque hacker conseguir extrair dos servidores esses dados, é o consumidor que será atingido.

3 Relação com outras disciplinas

A seguir teremos uma tabela relacionando a disciplina de Mineração na Web com outros campos da computação.

IF699- APRENDIZAGEM DE MAQUINA	Com os dados obtidos pela Web Mining, desenvolvedores podem otimizar seus sistemas com base no uso dos usuários.
IF706- TOPICOS AVANC.INTELEG.ARTIFICIAL	Sistemas de inteligência artificial geralmente precisam de muitos dados para criar um comportamento, então nesse caso é o Web Mining que provê.
IF782- NEGOCIOS ON LINE	O e-commerce sempre usa do Web Mining para receber dados de comportamento de possíveis compradores para o melhor direcionamento de propagandas e também para a otimização da experiência do usuário.
IF672- ALGORITMOS E ESTRUTURAS DE DADOS	Com os dados dos usuários de possíveis falhas no programa, os desenvolvedores podem usá-los para debugar e melhorar seus algoritmos.

Referências

- [1] L. B. Marinho and R. GIRARDI, “Mineração na web,” *CEP*, vol. 65080, p. 040, 2003.
- [2] P. R. Manning and H. Schuetze, *Introduction to Information Retrieval*, vol. 2008. Cambridge University Press, 2008.
- [3] P. F. B. e Ricardo Prudêncio, “Mineração na web,” 2018.
- [4] B.-Y. . Ribeiro-Neto, *Modern Information Retrieval*, vol. 2010. Addison-Wesley, 1999.
- [5] W. contributors, “Web mining — Wikipedia, the free encyclopedia,” 2018.