Mineração de Dados Web com Processos ETL em BASh Script

Vinícius Silva Madureira Pereira

bit.ly/spi2018b

Contexto

- O ensino da Mineração de dados
- Processos ETL
- Os benefícios das expressões regulares em relação ao custo do tempo de processamento
- Os constantes avanços do BASh

Vantagens

- Testes unitários a cada comando inserido
- Utilização de recursos do sistema operacional sem a necessidade de importação de bibliotecas
- Utilização de programas como recursos de linguagem
- Acesso à base de dados diretamente do *script*

Desvantagens

- Alto consumo de memória e processamento
- Conhecimento consideráve acerca de inúmeras ferramentas (programas e comandos)
- Não existe, ainda, uma padronização de desenvolvimento
- A Web é altamente dinâmica e o algoritmo pode depender de serviços de terceiros

Projeção

- Desenvolvimento de uma API para *scripts*
- Maturação de projetos em conjunto com aplicativos CMS
- Implementar, em conjunto com a Free Software Foudation (Projeto GNU) por meio da plataforma Savannah, uma biblioteca ETL

Referências

- JARGAS, A. M. Shell Script Profissional. 1 ed. São Paulo: Novatec, 2008.
- NEGUS, C. Linux a Bíblia. O Mais Abrangente e Definitivo Guia Sobre Linux. 1 ed. Rio de Janeiro: Alta Books, 2014.
- SILVA, L. C.; FERRAR, D. G; QUERO, P. Introdução à Mineração de Dados. Conceitos Básicos, Algoritmos e Aplicações. 1 ed. São Paulo: Saraiva, 2016.
- CEZAR, J. Programação Shell Linux. 10 ed. São Paulo: Brasport, 2014.

Referências

- MORIMOTO, C. E. kurumin 7. Guia Prático. 1 ed. Paraná: Sulina, 2007.
- PASSOS, E. Datamining. Conceitos, Técnicas, Algoritmos, Orientações e Aplicações. 2 ed. Rio de Janeiro: Campus, 2015.
- SEBESTA, R. W. Conceitos de linguagens de programação. 5. ed. Porto Alegre: Bookman, 2006.BEZERRA, E. Princípios de Análise e Projeto de Sistemas com UML. 2. ed. Rio de Janeiro: Elsevier, 2007. 369 p.

Dúvidas

Agradecimentos...

Muito obrigado!!!