

XXV Congresso de Iniciação Científica da Unicamp

18 a 20 Outubro Campinas | Brasil



Rastreador Web de Sinais Fracos

Lucas Miranda Tonon*, Ana Estela Antunes da Silva

Resumo

Proposta de desenvolvimento de um rastreador Web para a coleta de dados a serem analisados por uma ferramenta de inteligência antecipativa, visando apoiar a tomada de decisão empresarial.

Palavras-chave:

Inteligência antecipativa, Web Scraping, Extração de dados.

Introdução

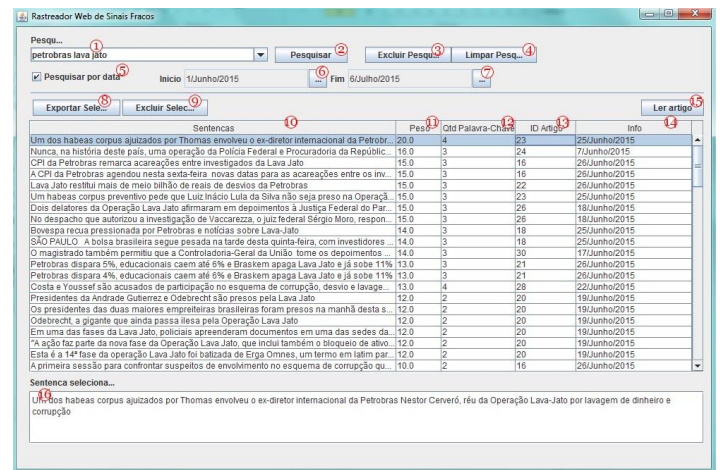
Com o fácil acesso à vasta fonte de dados e informação que é a Web, organizações vêm cada vez mais buscando informações ao seu redor para obterem vantagem competitiva. Porém, encontramos dificuldade em coletar, analisar e extrair informação útil, devido à gigantesca quantidade de informação encontrada na Web. Quando analisadas corretamente, o conhecimento adquirido por meio de sua análise gera vantagem competitiva para empresas, auxiliando no processo de tomada de decisão. Este projeto propõe o desenvolvimento de um rastreador Web, que será encarregado de cuidar do problema da coleta de texto na Web. Após os textos serem coletados, uma ferramenta já desenvolvida por um projeto Fapesp de Iniciação Científica (Processo: 2011/08696-9), orientado pela mesma professora orientadora do projeto corrente, SFIA (Sinais Fracos para Inteligência Antecipativa), analisará e extrairá informação útil dos textos coletados. A ferramenta SFIA é destinada ao apoio à tomada de decisão organizacional através de inteligência antecipativa. Esta ferramenta é baseada na tomada de decisões por meio da extração, armazenamento e análise de informações que antecipam acontecimentos e eventos, possibilitando a identificação de oportunidades estratégicas e ações antecipadas pelas organizações.

Resultados e Discussão

Como resultado principal este projeto obtivemos uma ferramenta que, a partir das palavras-chave escolhidas pelo usuário, realiza buscas na sessão de notícias do Google (news.google.com), visita os sites, extrai e mostra ao usuário as sentenças que contém as palavras-chave de cada artigo encontrado na busca.

Focamos, também, em estreitar as sentenças retornadas pela ferramenta. Antes, ela iria exibir todas as sentenças extraídas dos sites, e marcar as que continham palavras-chave. Desta forma era muito difícil encontrar algo útil, desde que o número de sentenças exibidas era muito grande. Para amenizar este problema começamos a descartar as sentenças que não continham nenhuma palavra-chave, limitar o número de sentenças de um site por 10, e criamos um ranking das sentenças.

Figura 1. Interface do software



Conclusões

A ferramenta desenvolvida cumpriu o objetivo proposto, possibilitando a busca automatizada de sentenças que possam ser sinais fracos, a partir de palavras-chave. O rastreador facilita a inteligência antecipativa automatizando a busca de sinais fracos. O usuário poderá encontrar uma quantidade muito maior de informação de qualidade, e em um tempo significativamente menor do que se ele fosse realizar as pesquisas manualmente. E a inclusão da função de pesquisar por data foi muito importante porque, para a inteligência antecipativa, uma informação pode ser irrelevante se não soubermos sua data da publicação.

Agradecimentos

Agradecimentos à orientadora Ana Estela Antunes da Silva por todo apoio e motivação.