

Web Information Mining and Risk Evaluation Platform for the Companies

ABSTRACT

This demonstration presents an web information mining platform WBRK. WBRK will provide the deep analysis about the risk evaluation about a enterprise. Enterprises can get helped from the analysis report about the latent cooperator or the competitor company. At the meanwhile, WBRK can also help the government to know the whole status about a company and make decision to support more daikuan or less. Currently, WBRK is applied to some big companies, most of them are shangshi companies.

1. INTRODUCTION

With the rapid development of internet, the amount of information is exploding. Millions of news report about different aspects of companies are generated everyday. In intuition, it's time to use those enormous information to support decision made by management espically on investment, merge, continue cooperation and so on. Form this point of view, the demand of web information mining and risk evaluation service is obversious for companies. Such services should include the following functions:

- a. The news reports of this system should be real-time updated and indexed.
- b. Analyzing the relationship between the news report and company.
- c. To generate short summary about the event related to company.
- d. To explore the employee trend about the compete companies.
- e. To evaluate the risk about the cooperator company.

In this demostration, we will present a web information mining and risk evaluation platomar, WBRK. The objective of WBRK is to provide risk evaluation analysis for both compa-

nies and government, including collecting web information, relationship analysis and provide risk evaluation. WBRK system focuses on the following functions:

- a. Relationship analysis functions is to identify the relationship among copanies. The relationship includes cooperation, competitor, such as A and B cooperate on somthing, A is competing B on something. These relationships can be detected from the news report through clustering algorithm.
- b. Summary the news reports around topic. The multi-document summarization method can be used to generate the summary, which help users to have a quick overview about the event without reading all the news.
- c. Evaluate the risk of specific company. We divided the risk into four parts, brand, opinions, employee trend.

Companies can get the whole analysis about the cooperator when new decision should be made.

2. SYSTEM DESCRIPTION

In this section, the architecture of WBRK and implementation of the core part are described in detail.

2.1 Architecture and Workflow

The WBRK system consists of three modules: realtime data acquisition, data analysis and generating the report. The architecture is shown in Figure 1.

With the help of data crawler, the real-time data are provided from the internet. The main data of this system is web news from big portal website such as sina.com, qq.com, etc. Then the system parses the raw web pages to get the main content and store them into local database for further processing.