Abstract

Existing search engines have many remarkable capabilities; but what is not among them is deduction capability—the capability to synthesize an answer to a query from bodies of information which reside in various parts of the World Wide Web. Web Intelligence is an area of research which attempts to provide this capability. 

Here in this paper we use Text Mining-a feature of Web Intelligence to derive information from the unstructured textual data on the web and device the consensus based strategy to business decisions. This will have two fold advantages, one mitigate the risk early and second would provide a support for our understanding and decision making. This concept is explained with an example of evaluating a player’s performance based on minute to minute commentary of the match. Parameters such as his position on field (for example in football – defenders,
midfielders, forwards and goal keeper), his past performance, his present fitness and form, and such other parameters are considered. Weightage / value for each parameter is decided and information can be derived for analyzing a player’s performance. During analysis we view the comments, we read through fan forums, blogs, newspaper reviews on the play, expert commentator views, etc. This is either used as a correction factor to enhance the credibility of the model.

The whole procedure involves four main stages: Web crawling i.e identifying information resources, information retrieval and extraction, text mining and finally converting unstructured data to structured data.

Reference

- GATE . www.gate.ac.uk. General Architecture for Text Engineering or GATE is a Java software toolkit originally developed at the University of Sheffield since 1995.
- Web Intelligence: kis.maebashi-it.ac.jp/wi01/ www.web-intelligence.com/
- Intelligence on the Web: www.fas.org/irp/irpintelwww.html WIN: home WEB INTELLIGENCE NETWORK, smarter.net/

Index Terms

Computer Science

Database

Management
Key words

Web Intelligence
NLP
Text Mining
Information Extraction
Information Retrieval
GATE