

# Business Intelligence

---

A dissertation submitted to The University of Manchester  
for the degree of Master of Science  
in the faculty of Engineering and Physical Sciences

2012

By

Mohib Ahmed Khan

School of Computer Science

# Abstract

The World Wide Web is expanding and will continue to do so. It has become the most preferred media for accumulating, exchanging and distributing information<sup>1</sup>. Companies and organizations today, offer a multitude of online services such as shopping, enquiries and support. As a result, they amass enormous amounts of data<sup>2</sup> in their day to day operations. If properly exploited, the raw data is a gold mine for information. The intense competition among different companies to acquire new customers and retain the existing ones has set the field of knowledge extraction from raw data into motion. This knowledge can help organizations devise competitive strategies, explore and benefit from potential markets, determine current and future trends and personalize the content of their customers.

Business Intelligence comprises of several techniques used to extract information of commercial significance from the raw data. The data collected on the web servers is termed as web logs. This data reflects a visitor's navigational activity, the time they spent on a page, their location and other related information. In this dissertation, web logs of a UK based company, Personal Audit Systems Limited are analysed. The company aims to sell their product, P11D organiser, an expenses and benefits reporting software within the United Kingdom. The project entails methods used to ensure that raw data is suitable for information extraction. It involves techniques used to extract valuable information. The project evaluates the role of many website pages and answers some company specific questions. Finally, it gives recommendations on certain areas of improvement of the website.

Keywords: Business Intelligence, Web Usage Mining, Web Logs, Data Cleaning

---

<sup>1</sup> information is useful knowledge derived from data

<sup>2</sup> data is raw, unanalysed facts, figures and events