"Analyzing two platforms .NET and Java for developing a search engine"

A Thesis
Submitted to the Department of Computer Science and Engineering
Of
BRAC University
By
Shoman Bhuiyan
ID: 02201046

In partial fulfillment of the
Requirements for the Degree
Of
Bachelor of Science in Computer Science
May 2008

Supervisor: Abdussamad Ahmed Muntahi (Lecturer)
Department of Computer Science and Engineering
BRAC University

BRAC UNIVERSITY

BRAC University, Dhaka, Bangladesh

Thesis Topic: Analyzing two platforms – .NET and Java for developing a search engine

Abstract:

There are several platforms available for developing a search engine. We can create search engine with php, Visual Basic, C#, J#, Java, DotNet and with other platforms as well. Among those, DotNet and Java are chosen for popularity. Both Java and DotNet use Apache lucene tools for search engine development. The criteria for this analysis were availability, readiness and depth of tools and features from user perspective. Java found to be a better choice and was used to develop a search engine for BRAC University website.

<u>INTRODUCTION</u>

A software that builds an index on text and answer queries using that index. Searching can be performed on local file system or during web surfing.
A **Web search engine** is a search engine designed to search for information on the World Wide Web. Information may consist of web pages, images and other types of files.
Some search engines also provides data available in newsgroups, databases, or open directories. Unlike Web directories, which are maintained by human editors, search engines operate algorithmically or are a mixture of algorithmic and human input.
A search engine offers scalability, relevance ranking, integrates different data sources (email, web pages, files, database and so on)

The very first tool used for searching on the Internet was Archie. It was created in 1990 by Alan Emtage, a student at McGill University in Montreal. The program downloaded the directory listings of all the files located on public anonymous FTP (File Transfer Protocol) sites, creating a searchable database of file names; however, Archie did not index the contents of these files.