

Faculty of Engineering and Science



Gooseveien 36, N-4890 GRIMSTAD

*Phone. 37 25 31 90 Fax 37 25 30
01*

Preliminary Project Report Web-Mining

Project: IKT407

Project 2 – Web Usage Mining

Written by :

Jørgen Andersen (jorgen@andersen.net)

Asle Ollestad (aslem04@student.hia.no)

Trude Buøy (tbuoy00@student.hia.no)

Date:

30.09.2004

PROJECT 2 – WEB USAGE MINING..... 1

0. SUMMARY..... 3

1. INTRODUCTION..... 3

1.1 BACKGROUND..... 3

1.2 PROJECT ORGANIZATION..... 3

..... 3

..... 3

2. GOALS AND EXTERNAL CONDITIONS..... 3

2.1 GOALS..... 3

2.2 ASSUMPTIONS..... 3

3. SOLUTION 3

4. PROGRESS PLAN.....3

5. REFERENCES..... 3

6. APPENDIX..... 4

6.1 PROJECT PLAN/ TIMELINE..... 4

0. Summary

Constructing a parser that segments a web log, identify the access sequences and evaluate the navigability of the web sites.

1. Introduction

1.1 Background

The purpose of this project is to make a parser that determines if a given web site is well organized.

1.2 Project organization

”Employer”:
Agder University College
↓
Supervisor:
Ole-Christoffer Granmo
↓
Project group:
Jørgen Andersen
Asle Ollestad
Trude Buøy

2. Goals and external conditions

2.1 Goals

The goal of this project is to construct a parser that segments a web log from a web site into sessions, and identifies the sequences web pages have been accessed within each session. Based on such sequences, traversal patterns of the web site are to be identified. Finally, it should be determined whether the navigability of a web site can be ranked meaningfully based on traversal patterns.

2.2 Assumptions

We have decided to use Python as implementation language.

3. Solution

The project consists of several tasks. The first step is to make a parser that reads a web log and extracts relevant information. This information is then sorted into sessions based on user id. Using data mining algorithms we will then analyze this data to find traversal patterns. These patterns will be used to determine whether the navigability of the website could be improved.

4. Progress plan

Phase 1: Learning to use Python

Phase 2: Web Parser programming

Phase 3: Evaluate program codes and functionality

Phase 4: Writing the main project report and prepare presentation.

5. References

Books:

Prentice Hall; Dunham, Margaret H.; Data Mining, Introductory and Advance Topics; 2003

Web pages:

<http://www.python.org>

6. Appendix

6.1 Project plan/ Timeline