CS 453: Project 1 - Crawling the Web

Project Report

Creepy - Web Crawler

Travis Hall
trvs.hll@gmail.com

Brittany Miller
miller317@gmail.com

Bhadresh Patel
bhadresh@wsu.edu

Washington State University Vancouver
September 12, 2010

Abstract

The main goal of this project is to design and implement a *web crawler*. The name of our web crawler is *Creepy*. *Creepy* is simple web crawler written in python that takes set of seeds (URLs) and start crawling the Web.

1 Overview

2 Automation

3 Document Storage

4 Politeness

5 Duplicate Detection

6 Roles

Travis Hall ...

Brittany Miller ...

Bhadresh Patel ...