NDSU DataSURG

Treeminer System Setup, Installation, and Usage Guide
Introduction

This is intended to be a living document for the setup, installation, and usage of the NDSU configuration of the Treeminer system developed by Mark Silverman of Treeminer, Inc., for use by NDSU students in vertical data mining research activities.

IMPORTANT: A non-disclosure agreement must be signed with Treeminer, Inc. before using the Treeminer system.

Treeminer specializes in vertical data mining solutions using structures known as pTrees. In simple terms, a pTree is a vertical bit slice consisting of ones and zeros. It represents values across multiple rows of data, with each bit corresponding to a single or partial value in its corresponding row. More information about pTrees can be found in Dr. Perrizo's presentations.

The Treeminer system is a robust pTree code base that includes methods for capturing raw data, creating pTree structures in Java, and various classification and clustering data mining operations using pTrees.

Quick Setup

1. Sign an NDA with Treeminer, Inc.
2. Request access to the NDSU pTree servers
3. Install VirtualBox
4. Download the preconfigured pTree virtual machine

Requirements and Setup

Getting started with the Treeminer code can be a bit of a daunting task as it is a very large and powerful system. To make the process easier for developers to start using everything, a few requirements and setup procedures should be completed before other steps are taken.

Non-Disclosure Agreement

First and foremost in using the Treeminer system is signing a Non-Disclosure Agreement (NDA) with Treeminer, Inc. Mark Silverman has been extremely gracious in allowing NDSU to work with the source code of his commercial software in order to facilitate collaboration.

In addition to working with commercial source code, Treeminer does business in industries and has relationships with clients where sensitive material and concepts are commonly addressed. This information must remain confidential and only shared between those who have signed the NDA.

To begin the process, an NDA document can be requested from Treeminer, Inc. through Dr. Perrizo. It is important to ensure that Dr. Perrizo is involved as the point of contact in order to validate legitimate NDA requests by NDSU students.

Access to NDSU pTree Servers

Once the NDA has been signed and approved by Treeminer, Inc., access to the NDSU pTree servers can be requested. Again, Dr. Perrizo must be involved in the request in order to ensure the NDA has been approved and the student should be allowed access to the ptree servers.

After Dr. Perrizo has given the go-ahead, the NDSU systems admin (currently Nate Olson) can grant access to the pTree servers. The username and password will be the same as what is used to sign in to the Linux Lab computers (typically first initial, last name; not the firstName.lastName login).
The pTree servers can be accessed through SSH (PuTTY) from anywhere, or secure FTP (FileZilla) on campus.

Virtual Machine Installation

The Treeminer system is implemented in Java and integrates with other open source technologies. It was originally created, and runs best, in a Linux environment. For ease of setup and to quickly get developers to a stage where they can start using the Treeminer system, a virtual machine has been created with the Treeminer system and many of its core dependencies installed and configured.

A virtual machine is a software solution to running any number of "computers" within an existing, running operating system. Using a virtual machine allows the Treeminer system to run in its native Linux environment regardless of the developer's primary operating system installed on their machine.

The virtual machine software chosen for running the Treeminer system is VirtualBox because it runs on all three of the primary operating systems (Windows, OS X, Linux). It is free open source software and can be downloaded from https://www.virtualbox.org/wiki/Downloads.

Preconfigured pTree Virtual Machine

After installing VirtualBox, the pTree virtual machine image can be downloaded from the pTree servers and imported into the VirtualBox list of virtual machines.

Once imported, the virtual machine can be started and will run like computer using a native Linux installation.

System Environment

The Treeminer system runs best in a Linux environment due to being written in Java. There can be unexpected quirks due to plugins and other dependencies that behave a little bit differently on Windows machines than on Linux.

Operating System

The operating system of the virtual machine is Ubuntu Linux. There are 32-bit and 64-bit versions available, depending on the developer's host machine capabilities.

Java

While the most current version of Java should work with the Treeminer system, it has been set up using a slightly older version to ensure compatibility.

Developers are free to experiment with updating to the latest Java JDK, and can reinstall the original virtual machine at any time to revert back to a fresh, working installation in case any problems arise.

Eclipse

The Treeminer code was originally developed using the Eclipse IDE. The project has been set up and configured in the pTree virtual machine.
Treeminer Code

The Treeminer project is very robust and includes a lot of existing code. It can be overwhelming when first starting with learning where everything is located and how to use it.

In general, one of the best ways to ease into the learning curve associated with using the system is to follow the structure of one of the existing classifiers.

**Important Folders**

TBD

**Important Classes**

TBD

**Important Tests**

TBD