# The Cyberscape Project

## Pushing Internet Browsing and Data Mining to a new level

Jason St-Cyr
#236901
95.495
August 31, 2002

# Abstract

This report addresses the issues, hurdles, and achievements that occurred during the development of the Cyberscape project. This project was developed in tandem with Ben Hall, and therefore for full understanding of the project, please consult Ben Hall's report for further explanations of server-side components.

This report addresses the client and user portions of the project. While time did not permit for implementation of all the necessary client features, the core of the client was built upon and discussions and design decisions were made to allow for future client features. The report will first discuss the goal of the project, and the deficiencies in current Internet technologies that provide Cyberscape with a place in today's and tomorrow's Internet community. Subsequent sections will explain exactly what the Cyberscape project aims to do, and which target audiences will be able to benefit from the project's applications. Throughout this process, discussions on issues and hurdles overcome will be highlighted and explained. In some places, suggestions for future overall development of general Internet Technologies will be detailed.

# Acknowledgements

This project could not have been accomplished without the joint efforts between myself and project teammate Ben Hall. Many of the design decisions, architecture, and base brainstorming was developed as a team, with development separated mostly by component (client vs. server).

Project Supervisor Michael Weiss also provided enthusiastic and supportive comments and discussions which allowed us to push in many directions we had not even considered before. Discussions with Prof. Weiss have provided us with a direction for future development of this application.

While this may not be entirely inappropriate, some acknowledgements must be provided to the developers of Sun's Java 1.4 which provided us with Regular Expression capabilities and XSLT processing power. We were also able to develop the entire project at low cost due to IBM's open source Eclipse project which provided a professional Java IDE from their website which works on multiple platforms with little error. Fellow Carleton students Paul Paquette, Shannon Borho, and Shawn French also helped me to develop a TFTP application in a previous term which was used to transfer files in our project. Without these resources, much of our project would not have been able to be realized in the time it took.

# Table of Contents

# 1. Introduction

The Internet. It's popularity truly jumped in the early 90's when the web browser was developed and home users could 'surf' to find information related to their interests. Some simply enjoyed the experience for itself, while others depended upon the internet to disseminate or find information. Still others found a new market for e-business and marketing. While technologies everywhere have improved to take advantage of the web, very little has been done to alter the way the web is experienced.

The Cyberscape Project was aimed at changing the viewing of the Internet in three distinct ways. The first was to change the 'paper and pen' two-dimensional viewing experience to a three-dimensional browsing experience. The second was to provide in this three-dimensional experience extra information about the internet which could not be seen in traditional browsers. The third was to try and enhance the sense of community when browsing the internet. The first few chapters discuss each of these goals.

The latter chapters focus on the methodologies of the Cyberscape project itself. This details development processes, design decisions, issues encountered during development, and other topics related to the development of the application. An overview of the final system as delivered is also provided, along with a short section on where application development will go next.

# 2. Internet Visualization: Analysis

There are several different users with different goals using different protocols to access the Internet. Tools are abundant to serve each groups needs, and for the purpose of HTTP, the traditional web browser has provided a means of accessing and providing data through the internet with an extremely simple interface. With the improvements made in the realms of Dynamic HTML (DHTML), Cascading Style Sheets (CSS), Flash, Shockwave, and other such efforts, the old boring pages of years gone by have been replaced by slicker versions of themselves. In the end, however, the interface has still not changed. You are presented with a flat version of the content, with few options for viewing meta-data about the site you are viewing. There is also no standard way of viewing the other users who are on the internet along with you. You are alone when browsing, unless the user decides to log into some sort of community in order to interact with others. These three deficiencies were identified as key areas of improvement for Internet browsing in general

## 2.1 Three-dimensional browsing

We exist in a three-dimensional world, however almost all of our technical advances still provide interfaces that emulate the two-dimensional 'pen-and-paper' methods of old. Most users never experience the possibility of a three-dimensional environment as they happily write their latest progress report in their favourite word processor. Why do we limit the interface to two-dimensions? It has been decades since the technology has been constructed to provide rich 3-dimensional environments, yet the technology has still not surfaced as a primary means of interfacing with information. What brings this about?

The problem lies not in the lack of technology. The technology is here, and is ready. The problem lies with the processes required to make that technology work. The first and largest hurdle is to somehow develop a three-dimensional interface which is intuitive and easy to use, yet disseminates all relevant data. Suddenly attempting to enforce a

new form of interface to a user, which does not conform to their previous experiences at all, will usually only lead to user frustration and lack of motivation to use the three-dimensional interface. The interface has to be readily understood with no training, and should not hide anything the user wishes to see. Designing such an interface is extremely complex, and in most cases more costly then the benefits.

The second hurdle lies in the hands of those who control the finances. The cost of developing a three-dimensional interface is much higher than for developing a simple two-dimensional interface which can work off of pre-existing technogolies. The current interfaces are quick, well-understood, and do not seem to have any large drawbacks. A customer or project manager is likely to ask: “Why do I need a three-dimensional file-manager, I can just drag and drop and everything works fine”.

In the end, it becomes apparent that most of the computer technologies developed do not require a three-dimensional interface to improve their performance. Most of the software in common use, such as word processors, spreadsheet programs, graphics programs, are oriented around ‘pen-and-paper’ uses, and therefore a ‘pen-and-paper’ interface is the most appropriate. If the goal of the software is to automate or provide a digital interface to something normally done on pen and paper, then of course the interface should match. However, there are certain technologies which would be much more appropriate in three dimensions.

Imagine if you will a hardware troubleshooter that provides a three-dimensional imaging of the internals of your computer. As a network technician, you can enter inside the box without ever opening it and view the contents and make ‘repairs’ as needed. Configuration for adding new hardware drivers, etc. could be easily improved with such a tool.

Another example is software for house-planning. This software has already moved to a three-dimensional interface. Architecture, interior design, construction, etc. can all

benefit from the use of three-dimensional modeling to mimic a three-dimensional environment.

However, when viewing web content, the chosen medium is a two-dimensional representation of the browsing experience. Regardless of the fact that users traverse through the world on digital transmissions, and interact with people all over the Earth, a two-dimensional showing is all that is provided. Our project begged to ask why this should be, and determined to lay the foundation for a three-dimensional browsing experience.

## 2.2 Internet meta-data

There is a wealth of information about the usage of the internet that is not readily available to the average user. It is impossible, for instance, to know which sites a particular website is linking to, or know how many visitors are at the website at that current time. Many sites which attempt to provide a community will add counters and 'Links' sections and login facilities so that users can see who is logged on at the moment. Often times, however, the links section is simply a section of important links. The login facilities require the user to be a member of the website, and the counters rarely provide information that is relevant to the user.

Recognizing these deficiencies, the Cyberscape project aimed to overcome these issues by providing a level of abstraction away from the site-specific meta-data. The data that really concerns a user is the comparison between different sites. If one site has 300 hits, and another has 500 hits, then the user can determine relative popularity by the data provided. Regardless of what the number actually represents, due to its standardization across the Cyberscape project the relative values are still valid. The project was also developed to allow for the tracking of relations between sites, which would provide users with a way of viewing how sites are linked to each other, and which

links are used by individuals more often. The last issue Cyberscape tried to address in this field was meta-data regarding other users in order to enhance the community approach. Design decisions were done for the project to allow for the eventual integration of multiple-user avatars, much like some of the three-dimensional chat worlds that have been created already. The Cyberscape project should allow users to be able to view metadata about other users in relation to themselves, and also provide information as to how many users of the Cyberscape are in the local area. Much of the meta-data implementation was organized around the visualization of popularity for the time being, with community and links implementations slotted for development beyond the scope for the Honours Project deadline.

## 2.3 The Internet Community

Related to the idea of multiple-user avatars in the environment was Cyberscape's attempt to work on the community aspect of the Internet. From chat rooms/worlds, to IRC and instant messaging programs, the ability to interact with others through the internet has been a long-time favourite pastime for users. However, this idea of community and personal interaction has always been separated from the web browsing experience. Due to the nature of the traditional display, there was no way to conveniently provide all the tools for messaging within your browsing window. Most of the time, it would simply be an interference.

In an attempt to provide this community, the Cyberscape project was designed to allow multiple users to interact with each other. Using a 'chat-world' approach, the browsing of the internet information is integrated into the 3-dimensional world where inter-personal communication can also take place between users in their avatars. Not only would this provide for an actual sense of community while browsing, it would also allow you to interact with others who are accessing the same information at the same time. The possibilities for this are nearly endless, and would provide utilities for several different target groups.

# 3. The Cyberscape concept

This chapter will provide a more detailed overview of the Cyberscape project's envisioned utility. Though much of the project's capabilities are still in a theoretical phase, an examination of the various components and how they are used is essential. Of foremost importance, however, is explaining why this project was conceived, and who will be able to benefit from it. In the previous chapter we discussed some of the deficiencies in the current browsing experience, and now we will examine these issues further and illustrate other areas where the Cyberscape will prove to be a useful tool.

## 3.1 Uses and Applications

**Why build the Cyberscape?**

Currently, the Internet is a collection of servers and websites that have no real absolute link between them. There are hyperlinks on the websites themselves to try to offer paths between certain sites. There are search engines that can group sites by search criteria. There are advertisement banners that provide a means of linking from one area to other areas of the web. There are sites that offer directories of links to certain types of websites.

However, these links are arbitrary and website dependent. There is no sense of a real structure or world behind these websites. The concept of the Cyberscape community proposed is to provide this virtual world that lies behind all these websites, providing the user with an interactive three-dimensional landscape in which to experience this world.

This virtual world is meant to be capable of using visual metaphors to display the relations between websites, the size of a website, the popularity of a website, paths traveled by other users, other users in the system themselves, and other information which may be desired.

**Who would be interested in the Cyberscape project?**

We have noticed an unending supply of venues which the Cyberscape project can encompass. At the moment, with the scopes envisioned, we have identified four key target fields in which the Cyberscape application would be of use:

**1.** *Web advertising:*

Companies marketing or advertising on the web wish to be able to maximize their revenue from paid ad placements on websites. Cyberscape provides these advertising companies with the ability to visually locate and examine the sites which have the most traffic and to see which sites link to each other so as to maximize coverage while minimizing overlap.

**2.** *Web data mining:*

Many companies are in the business of mining data about people's web usage, mostly about the websites they visit. To be able to function and build its three-dimensional environment, Cyberscape needs to track much of the same data these data mining companies acquire. However, Cyberscape is able to provide a three-dimensional interface to the data, which most data mining companies cannot. The Cyberscape concept will allow these companies a way of visualizing the data they have collected. Also, due to the way XSLT has been implemented to determine output, the client can easily add their own stylesheets to conform the data to the format they would like for reports or printouts.

**3.** *Internet Service Providers / Static IP Providers:*

Several large communications corporations (Bell, Rogers, Magma) offer IP addresses to their cable and DSL internet customers. Often, in the Terms of Service, these companies request that their users not install webservers or run websites on this IP. The Cyberscape tool uses a visual metaphor to display how many domains are associated to an IP/ This will allow these companies to locate any users who may be

violating their Terms of Service. The popularity visualization will also allow these companies to locate users who may be stressing the system. In short, the Cyberscape provides the owners of an IP a way of monitoring their IPs using visualizations of IP and Domain data.

**4.** *Web surfing / Instant Messaging*

Users of the web tend to fall into a few distinct categories, one of which we have coined the 'web surfer'. This person will randomly browse through the net looking for something of interest, or perhaps they simply find enjoyment in browsing the net. The Cyberscape community will allow these web surfers a way of finding websites that are extremely popular without having to search through millions of other websites. The community aspect will also allow these types of users to congregate while surfing, potentially travelling together as they continue their surfing.

## 3.2 Cyberscape components/features

This section will not discuss architectural components, but rather the components and features that would be of interest to the user. Some of these features were not developed in the demo software, but were nonetheless part of the project's design and architecture phase. Further chapters will discuss reasons why certain features were not implemented for the demo.

**1. 3D Browsing**

The initial concept of the Cyberscape community was to offer a three-dimensional browsing utility. This would be complete with full translation of HTML websites into three-dimensional objects on the users screen, as well as offering a way to leave a website and then continue travel through the Cyberscape to other websites. Also, travel between areas of the web was supposed to be made possible by having various levels

of browsing abstraction. The base abstraction would be at the website itself, viewing the content in three dimensions. The second tier of abstraction would be the IP address building itself, which housed the various domains at the IP. The third tier of abstraction was termed the 'neighbourhood', which contained a three-dimensional cityscape representation of the various IP addresses surrounding the user. The fourth tier of abstraction, termed 'cities', offered ways of navigating between 'neighbourhoods'. The fifth and final tier allowed for the navigation between 'cities'. At each level, meta-data would be able to be viewed at the scope of view chosen.

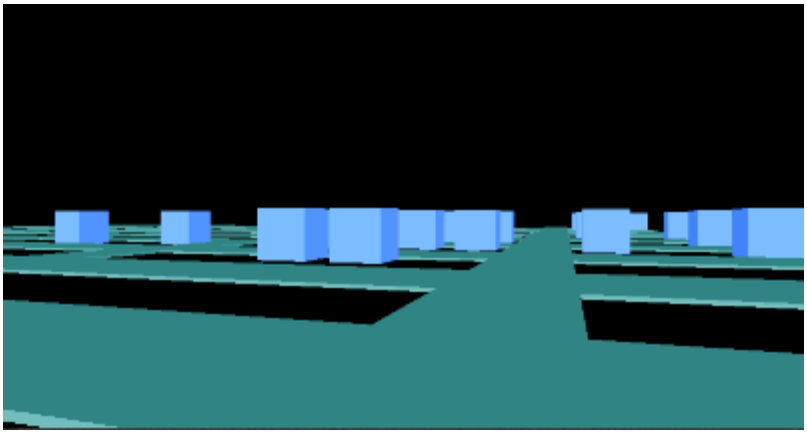

**Figure 1: VRML representation of internet [msn.com requested]**

**2. Popularity data visualization**

Possibility the most integral portion of the system was the tracking of web usage. Without this data, all objects in the three-dimensional landscape appear the same, with no distinguishing traits. Using visual metaphors, this data is represented to the user in a relational way. As mentioned before, the comparison of 'hits' between websites is more valuable then meaningless numbers displayed on individual websites. Using this as a base, the popularity data is viewed in relation to each other. Perhaps the site did not actually have 3000 hits, but as long as all sites are being monitored in the same way, the relational data is still relevant. Since a 'cityscape' visualization was used for the project, the visual metaphors employed have followed this concept. Popular IP Addresses should appear as skyscrapers, reaching upwards, while other less popular buildings would be shorter and less noticeable in the landscape. The number of domains held by

an IP Address should be shown with varying qualities of textures and materials for the buildings, but has been shown with varying colours in the demo software.

All of the 'buildings' (IP Addresses) were also associated with streets where the user could walk down and view the various internet sites. Popularity data on these street segments were also tracked so that other users could view which portions of streets were more popular. This would allow for users to watch common paths develop in the Cyberscape world. In the demo software, street segment popularity was shown using varying colours.

**3. Location to location transportation**

Standard web browsers offer the user the ability to enter a specific location to go to, rather than following links from their homepage. This feature was also desired for the Cyberscape project to allow the user to specify a specific location which they would like to view. By entering a domain name or IP Address, the client requests a map of that location and is transported in their browser to that map. This allows the user to navigate the Cyberscape with direct location to location transportation, rather than always following the streets.

**4. Community software.**

The Cyberscape project's initial design phases also included the desire to bring a sense of community to the browser. Initial design called for multiple-user avatars with chatting capabilities and preference-based visibility in the 3-dimensional world. Development to this point has been to allow for this eventual possibility, though not all of these features could be implemented in the scope of this deliverable. An IRC client was added to the software to allow the users to communicate with each other, even if they could not see each other in the virtual world. Eventually, a three-dimensional chat-room capability should become part of the landscape, allowing users to browse the internet while interacting with other users of the application. Specific details concerning the IRC client are available in Ben Hall's submitted document.

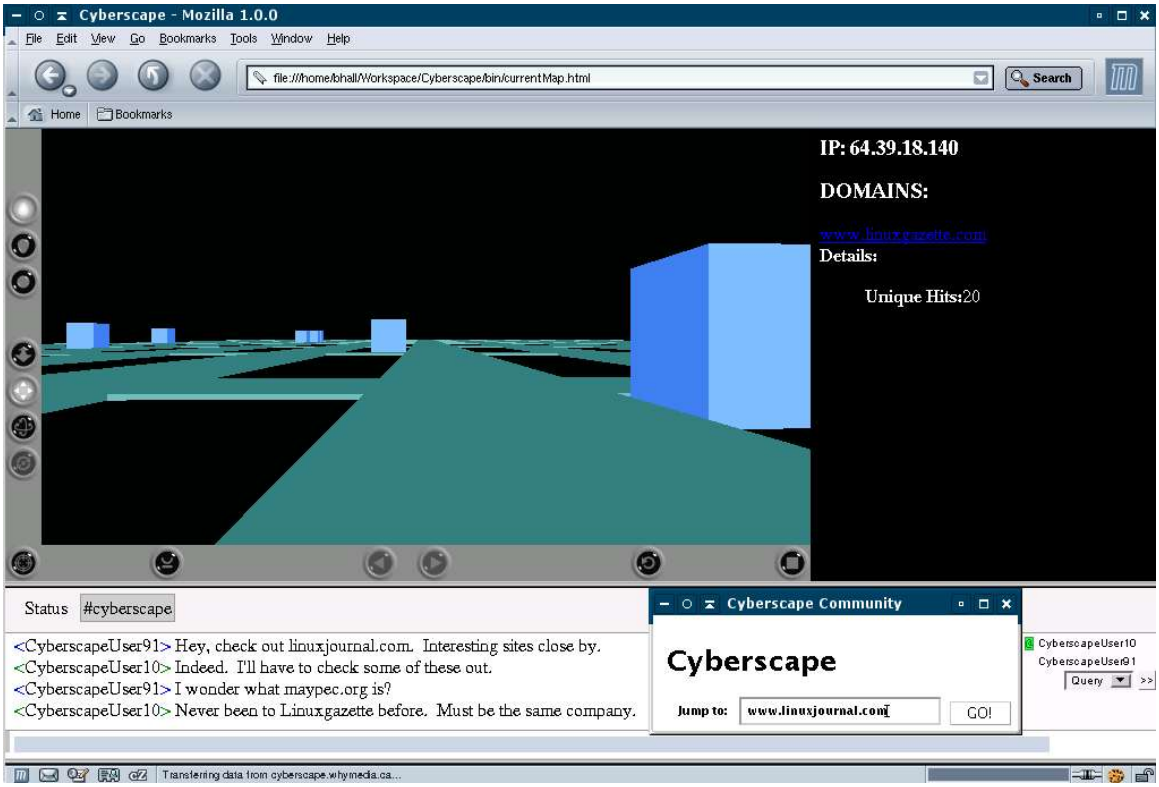


**Figure 2: The Cyberscape Client illustrating the Integrated IRC Program**

### 5. Global Positioning System

Related to the display of popularity data was a GPS system, which would have provided a 2-dimensional overview of a large area, allowing the user to view detail about the map in which they were currently located. This would allow users to find where they are in relation to other popular elements, and allow them to view popular paths near their current location. While not implemented for the first iteration of this product, the system's design has allowed for the addition of this feature at a later date.

### 6. Client web usage tracking

To gather all the important data required to display the information to the client, the initial system design called for a method of tracking client web usage. For the demo software submitted, this was accomplished through the use of an HTTP Proxy which tracks and submits locations visited by the client. More details on this feature are available in Ben Hall's submitted document.

# 4. The Cyberscape implementation

As part of the Honours project, the core of the Cyberscape application was designed and implemented to explore the possible barriers to this type of technology, as well as show an example of where this application could be taken in the future. Over these few months, several problems occurred, but almost all were solved in one way or another. The next few sections will detail the various parts of the implementation that this student was involved in directly, and some information will be provided on relevant problems encountered by the other student on the project.

## 4.1 Transferring gathered data through TFTP

An implementation of the Trivial File Transfer Protocol (TFTP) has been added to the Cyberscape project to allow for the transfer of files between the central server and the various clients. The project required a very simple way to transfer files between a client and a server, without having to develop a new transfer protocol and software from scratch. Having already developed a TFTP application for another course, it seemed logical to add the application into the Cyberscape project files to handle our file transfer. However, the implementation of the TFTP application as it was would not allow us to silently run it as a shell program. The application was improved to allow for silent running from another application using command line parameters. This allows the Cyberscape applications to send files back and forth using packet verification and validation, as well as verification of duplicate packets.

When attempting to launch the TFTP server developed, it became readily apparent that there would be a security issue with the software. The TFTP Protocol requires the server run on Port 69, which requires root access to the server to run. This would open a potentially dangerous venue onto the Cyberscape server, since no security measures (such as login/password/encryption) is used in the TFTP implementation. To circumvent

this issue, the port used for the Cyberscape project is not that specified by the TFTP protocol, allowing the server to be run by a non-root user.

File transfer was chosen above messaging due to the large amounts of data that need to be transferred. This particular TFTP implementation, unfortunately, is an extremely slow transferring system relying on UDP and only ever sending one packet of information at any given time. Future implementations of the Cyberscape project might use more advanced transferring methods, but at this time the use of a readily-developed TFTP system reduces development time and effort.

Aside from the port issue, the TFTP application used follows the TFTP protocol specifications, which can be downloaded from the following websites:

STD0033: ftp://ftp.isi.edu/in-notes/std/std33.txt

RFC1350: ftp://ftp.isi.edu/in-notes/rfc1350.txt

## 4.2 Cyberscape data in XML

Design decisions led to the use of XML for transmitting the data on the Cyberscape world to the client. The XML format allowed the client to implement a standard XML parser which provided data parsing and searching without the need for our group to develop the parser or search engine. Also, the use of XML allowed for the use of XSLT in transforming the data into necessary output files. This also led to the ability to rapidly change how the client-side display is done without altering the code used to generate the XML data files.

XML has provided the opportunity to provide a structure to the data being transmitted from the database, and has made the implementation of the system much easier by reducing the development work required. However, by eliminating the need for a parser or search engine, the work was replaced by having to develop the application base to support the generation of the XML data file.

First, an XML schema had to be designed to support all the information necessary for the Cyberscape application. As development proceeded, several rewrites to this format were required as it was realized that more refined specifications were necessary. Once the format was developed, example XML files were created by hand to test the XSLT stylesheets used to generate the Cyberscape environment. Having developed a format which adequately provided the data required to generate the environment, the next step was to develop the builder code which would create the XML file. This code had been foreseen as a small matter, however once development began, it became quite obvious that there were several problems which would require many hours of design and development. Eventually, this student managed to design a system of builder objects which would know how to construct themselves into an appropriately structured XML file. While the structure tended to be fairly predictable, the data was not. Depending on the user's location in the system, different maps were required to be loaded, and therefore the data in the XML files had to be dynamically retrieved from a storage space. This led to the final step in the chain which was the design and implementation of a database to hold all the relevant data about the Cyberscape world. This part of the project was handled by Project partner Ben Hall, who developed all of the tables, as well as the Java wrapper code to allow the XML builder objects to access the database and retrieve the information necessary.

## 4.3 XSLT translation of data

By using the powers of XSLT, the application is capable of translating information from one format to another. The first goal had been to translate XML data about the Cyberscape world into VRML that can be viewed by the user in their web browser (using a VRML plugin). The second goal was to use an XSL stylesheet to be able to translate the XML data into detailed HTML files summarizing domains available at an IP address. During the design phase, the possibility of using XSLT to translate HTML into VRML or XML was also raised This would enable interfacing with web search engines to translate

the HTML results into viewable and transmittable data. This design issue was not implemented, as there was little time to implement a search engine interface.

The XSLT translation was accomplished using Sun's JAXP package (**J**ava™ **A**PI for **X**ML **P**rocessing) provided with Sun's JDK 1.4. Using the API provided, a simple transformation can be done by providing the input file, the XSL stylesheet to be used for the transformation, and a location for the output.

The XSL stylesheets being used have been developed in a navigational style, meaning that it expects a certain structure to be in place. At a later date, a rule-based implementation may be developed to allow more flexibility from the input file. However, at this time, the navigational style was more simple for development and therefore allowed for quicker results.

Preoiminary development on the XSLT steylsehets was done using one central stylesheet which was capable of including other stylesheets which could translate the data to specific formats. For the current scope, this included one stylesheet for VRML and one stylesheet for HTML. Preliminary tests and runthroughs seemed positive, allowing for a hierarchy of stylesheets to be used. This would allow users to develop their own steylsheets and add them to be included in the central stylesheet to allow them to view data according to their needs. It would also allow for easier updates to client-side code by only having to devlop new stylesheets for new features required. However, the JAXP Xalan parser used in development would not execute the stylesheets in order, resulting in mixed output. The parser also lacked the ability to perform muliple document outputs, which forced the development of file splitting utiltiies using JDK1.4's Regular Expressions packages. In the end, some of the main features hoped to be brought to the project by the use of XSLT were lost because the JAXP XSLT transfomer could not fully support the XSLT 1.1 specification. Future iterations hope to use an improved XSLT parser to achieve the missing funcationlity and remove the hardcoded filesplitting technology.

## 4.4 VRML mapping of the Cyberscape environment

For the scope of this project, VRML has been chosen as the output language for viewing the data in a 3-dimensional way. Visual metaphors are being used to represent data from a distance, providing the user with the ability to quickly discern relative information about websites and IP addresses. Additional textual data in XHTML is provided in a different frame to allow the user to view textual details on selected targets more clearly.

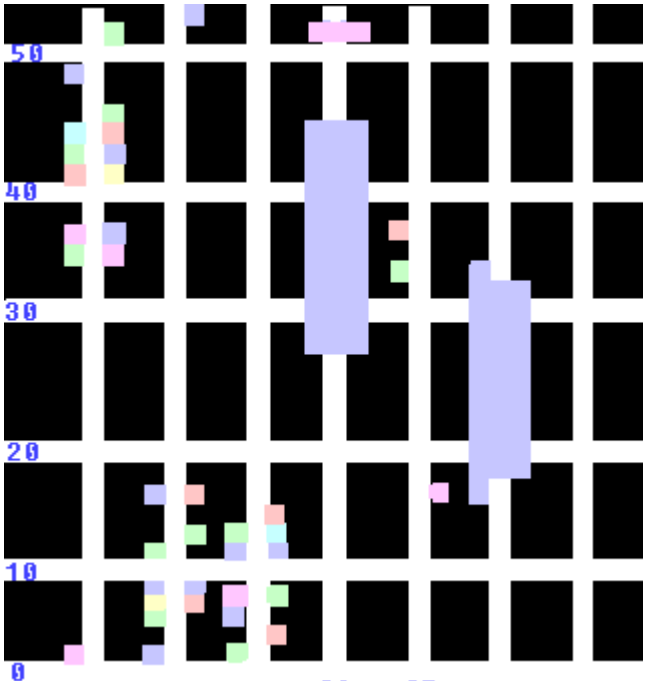


**Figure 3: 2D overview of IP-based street map**

Currently, the mapping representation being used is based on IP addresses. Streets and neighbourhoods and buildings are all displayed based on IP, and the IPs popularity. Adjacent IP Addresses all exhibiting the same domain name are planned to be visually grouped together into connected buildings, although this level of complexity has not been added to the application for this iteration. Streets are provided to allow the user to navigate, but also to show popularity of paths. Each street segment has its own visual representation of its popularity, allowing users to find pathways which they are more likely to be interested in. The image in **Figure 1** provides an overview of how the 207.219 ‘neighbourhood’ might appear as seen from above.

**What is IP popularity?**

Most of the visual metaphors rely on usage, hits, etc. The more times a user visits a domain at an IP, the more popular that IP becomes. In result, the building grows higher as it reaches through the various boundaries of popularity to reach a new level. The building height represents the IP Address popularity, but the colour represents the number of domains the IP has to reach this popularity. If an IP has many domains and is tall, this is visually shown, and can be compared to an IP that has the same height but has a different colour because it has fewer domains.

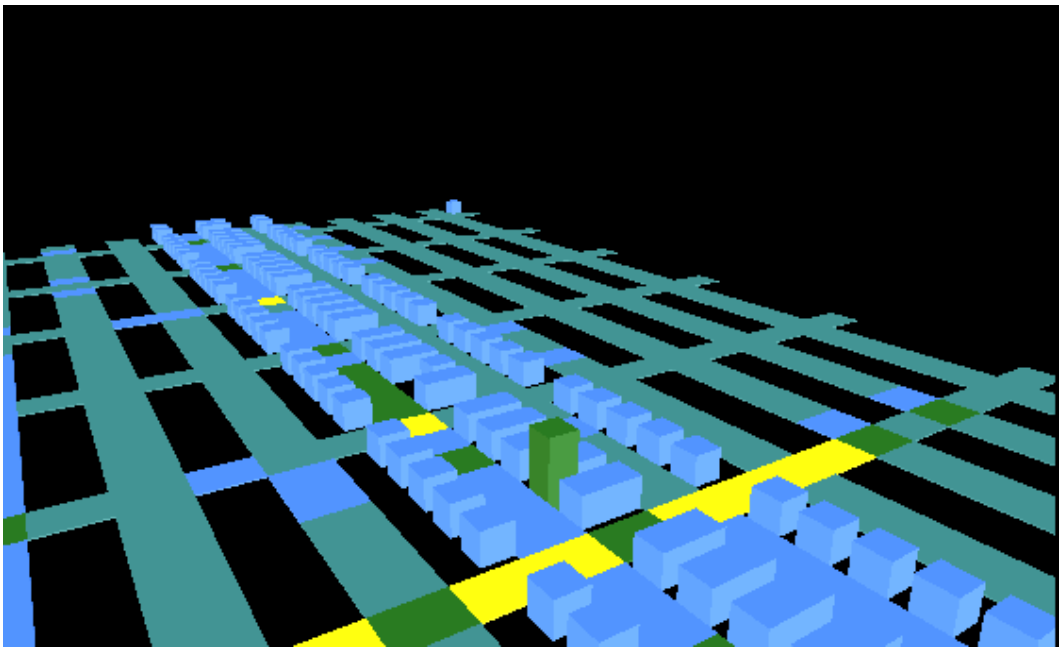

**Figure 4: VRML map showing various popularity metadata**

While this is the ideal design, actual implementation had to be altered somewhat for the demo provided. When it was discovered that domain names can be associated to multiple IP addresses, a small problem arose with tracking of IP popularity. Since most users do not enter the IP address of the site they wish to visit, there was no way for the Cyberscape application to track the IP-specific popularity. Popularity would then have to be ruled by domains. For more information, see the **What is Domain popularity?** section further on.

**What is Street popularity?**

Every time a user in the VRML world passes over a specific segment of street, the popularity of that street segment rises. In future iterations, when a user transports directly from one domain to another, changing IP addresses, the shortest path between the IP addresses should be calculated and the street segments along the path will increase in popularity. This design decision did not make the list of top priorities for development, as it did not display any extra power of the application.

**What is Domain popularity?**

Up until this point, we have not spoken about domains in the Cyberscape environment. Each domain's popularity is also tracked by the system, though it is not immediately evident. When a user visits an IP, they can break down the IPs popularity down into domain popularity. Domain popularity represents user's visits to a specific domain name. If a user requests to go to a specific domain name in the VRML world, this increases the popularity of the domain (and all the IPs associated with the domain name). When a user is traversing the internet, the client-side proxy tracks their usage and reports the domains they have visited to also track popularity of domain names. Domain popularity is shown as special content, and is not available through the regular street map world shown to the user. At a particular IP, the user requests domain popularity breakdowns of that IP. Future implementations will also allow for area sweeps for most popular domain names, etc.

This particular implementation for domain popularity tracking, as is probably obvious, has some flaws. If a domain visit updates all of the domain's IP addresses with a hit, then the IP popularity will be skewed. This design decision was made since no simple way could be found to determine which particular IP is being visited by the user at that domain name. As a result, individual IP address tracking had to be dropped, and an IP addresses popularity is now the sum of its domains popularities.

## 4.5 Map Request system using Socket transmission

Based upon the architecture developed for the TFTP application, a similar system was developed to transmit simple single-packet map requests to the server, waiting for an eventual response from the server. The client needs to transmit the current location of the user to the server, which receives the request and begins creating the XML map file for the client. When the map has been constructed, the server responds to the client with an HTTP URL to the XML map. Currently, there is no way that the user can cause this sequence of events to occur from within the browser. Ideally, when the user moves off the current map, the browser should launch a script which will cause the application to communicate the map request to the server. Currently, new maps can only be requested by the user's requests through the Cyberscape Java GUI provided (see section 4.6). To properly implement the ideal behaviour, it is likely that a closer integration of the application be made with the user's browser to allow for complete communication between the components.

## 4.6 Java GUI Console

The entire client-side application begins with the Java GUI Console. The console is very simple, providing a field for the user to enter a domain name or IP address that they would like to view. The console then passes off functionality to a handler object which performs all necessary steps by executing various Cyberscape components. The handler requests an XML map, converts it to VRML and HTML, and then launches the user's browser onto the VRML world representation of the location they requested. During the design phase, it was decided this form of interface would be the simplest to demonstrate the funcationlity without having to spend development and design time on researching browser integration with the console. Ideally, this sort of interaction should be completely within one interface so that map requests can be done from the same browser which displays the VRML world. The simple GUI console has been designed to be able to be replaced by another interface with minimal impact on the system.

# 5. Cyberscape and the Internet: A look forward

This document has already explored certain venues of development that the Cyberscape project will take in years to come, however this section will discuss in general where the project is aimed to proceed, and where changes to the current structure of the internet will be required to take us to this level. With time, this project will be able to be widely used, and will hopefully show users a new way of using their online time.

## 5.1 The future of the Cyberscape project

With the completion of some of the core functionality required for the project, work can now be done on bringing the project to an acceptable level of useability. The 3-dimensional content will be improved to be able to handle multiple user avatars, as well as improved display materials. Loading of adjacent maps will be a prime concern, since this will allow users to walk from one area to another. Integration with search engine functionality and display of websites in three dimensions as well as 2-dimensions will allow the user to stay within the Cyberscape world when viewing websites.

Some avenues of exploration will include vehicular transport, public transportation, public areas of congregation, user preferences, the GPS system (as discussed in Chapter 4) and other community features to enhance the users experience while within the 3-dimensional world. The goal is to bring web browsing, messaging, and other community features together into one package.

While data-collection was mostly discussed in partner Ben Hall's document, this is an important piece of the Cyberscape world, and is a huge avenue of advancement. Currently, only popularity data is collected, but relational data based on links to other sites would also be of interest, as well as visually displaying links between sites in the browser. Additionaly, other new forms of data would be relevant to track, such as length of time that domain names have been in use, or data on current visitors to a certain

location. This type of data would allow users to determine if some sites are less popular because they are newer and haven't garnered as many hits. It would also allow users to view which locations are more popular at the moment, as opposed to historically.

Another change that will need to be looked at is more specific tracking of web usage. Currently, the system only tracks total hits, and does not specify the time over which these hits occur. Yearly, monthly, daily, hourly breakdowns would be helpful options for the user in displaying the world. For example, the user could change their preferences from 'total' to 'daily' display of hits, and the data could be translated into a new map which would then be displayed according to the user's request. This type of translation power can be provided by XSLT in a rather simple fashion.

One other avenue to explore is increased data tracking. Currently, the system uses an HTTP proxy to track user usage, but without widespread use of the proxy, the data in the Cyberscape world is skewed. New ways of gathering web usage must be explored in order to offer a Cyberscape display which accurately represents the internet user population's usage.

While brainstorming, some rather interesting ideas came to mind concerning possibilities of taking advantage of the current interests in web-based media and file-sharing. There are moral and legal issues to deal with in this realm, but some of the ideas that have been talked about include adding music to the 3D world. This would allow streaming audio radio stations that the user can listen to as they walk around the Cyberscape. The other idea involved the development of 3-dimensional 'homespaces' which would be offline client-side 3D representations of their file-system, allowing them to use the Cyberscape's 3D rendering capabilities to view their system, and choose which files to share online with others. This would open up avenues for social gatherings online at user's 'homespaces' where file-sharing and online chatting can occur. All of these are simply ideas at this point in the design process, but attempts to design the system to

allow for these possibilities have been made.

## 5.2 Improving the Internet to meet Cyberscape requirements

One of the largest issues encountered when developing this software was the inability to request a list of all domain names. The only way to be able to completely construct a world with all the domain names available would be to have a list of these domains and then query for their IP addresses. However, it was quickly learned that such a task was impossible. To work around this, several tactics were employed, but it would have been much easier if the DNS server could have been queried for a list of all domain names. The processing power would have been exhaustive, but would have been far more efficient then the methods employed by our team. This was one of the main reasons why construction of the world's had to be based on IP addresses, because the IP address was the only constant that could be found.

Another large problem with the current technology available is the lack of standardization amongst web browsers in displaying and rendering content. Even though the W3C has set forward a ruleset for HTML, browsers handle pages differently, and different people interpret the standard in different ways. This results in non-standard HTML documents that can be formatted in several different ways but provide the same output. When attempting to find a way to translate HTML into a 3-dimensional view, the sheer task of parsing and deciding what to do is enormous. A stricter standard would allow for parsers to more readily handle HTML documents for conversion. Another problem is web enhancements such as Flash and Shockwave which offer content without adhering to any HTML standard whatsoever. This type of content cannot be altered to 3-dimensions without considerable effort, and likely would have to appear as-is with the use of a plug-in in the Cyberscape browser.

A third improvement which the Cyberscape team would like to see is an ability to distribute the data collection and storage used by the Cyberscape's view of the internet.

Much as DNS servers have been distributed to serve current Internet browsing needs, a similar system would be extremely helpful in supporting the Cyberscape data. By farming out collection, storage, and map request serving to different servers the entire application would benefit from a huge increase in efficiency. Currently, the system relies on all users contacting a central system which handles all requests, collects all data, serves all data, and constructs all maps.

# 6. Conclusion

Over the last 4 months, many man-hours were invested in the development of this initial iteration of the Cyberscape project. By taking advantage of other technologies that have already been developed and integrating them together, an extremely large application was able to be constructed in much less time than would normally be required for such an undertaking. The ability to re-use and fuse pieces of other applications for our benefits was one of the key learning experiences of the project, and allowed us to see how otherwise unrelated tools could be used to join together and create something new.

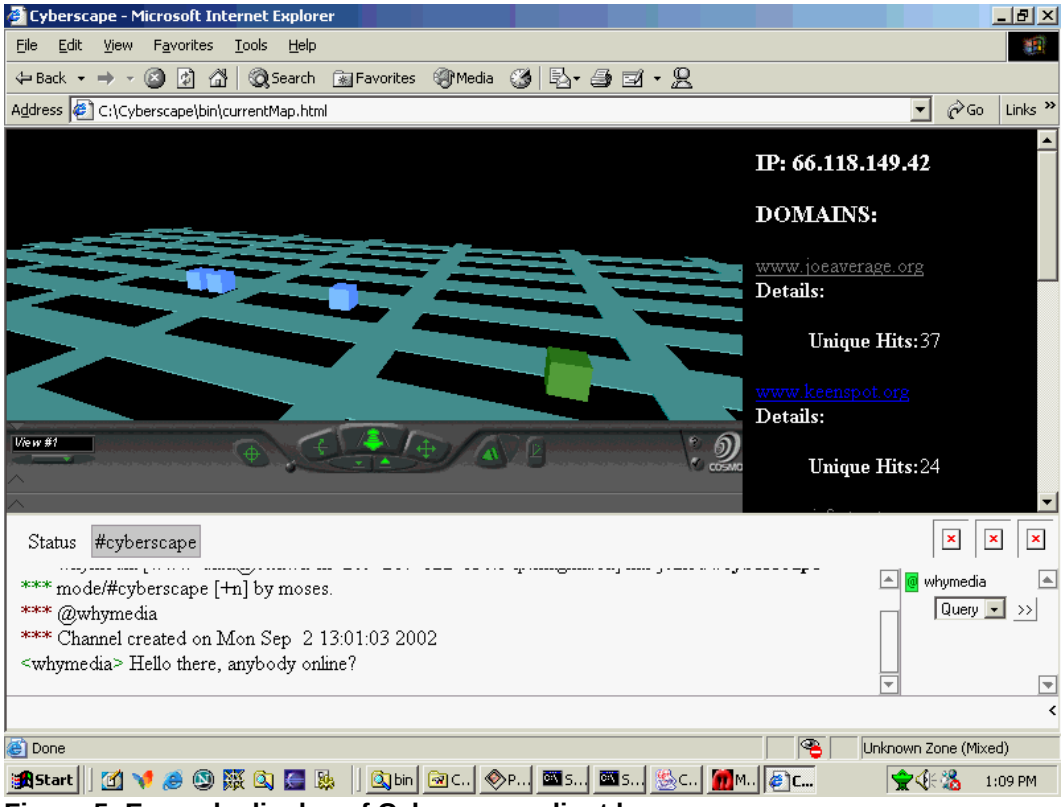

**Figure 5: Example display of Cyberscape client browser**

The team encountered many issues, but with access to the wealth of information available on the internet all issues were solved in relative short order. While not all of the features desired in the end-product were able to be developed, the Cyberscape project can definitely be seen as a success. The team was able to display the ability to

develop the core functionality and prove that this sort of system will work, given time and resources. My personal forays into VRML, XSLT, XML, and Java provided me with a wealth of learning experiences and new ways of looking at each language. I was also able to realize just how limited the current technologies are, and can better appreciate where these technologies will evolve in the future.

# Appendix A: XML Map format

The following is the Document Type Definition (DTD) for the XML format for transferring information about a specific map to be loaded by the client. All XML files generated from the database must comply with this DTD.

**DTD:**

```
<!-- Map Elements -->
<!ELEMENT Map (Street+)>
<!ATTLIST Map id CDATA #REQUIRED>
<!ELEMENT Street (StreetSegment+, Intersections+)>
<!ATTLIST Street id CDATA #REQUIRED>
<!Element StreetSegment (Geometry, Info, InetAddresses)>
<!ATTLIST StreetSegment id CDATA #REQUIRED>
<!ELEMENT Intersections (IntersectionSegment+)>
<!ELEMENT IntersectionSegment (Geometry, Info)>
<!ATTLIST IntersectionSegment id CDATA #REQUIRED>


<!-- InetAddresses Elements -->
<!ELEMENT InetAddresses (InetAddress*)>
<!ELEMENT InetAddress (Geometry, Info, Domains)>
<!ATTLIST InetAddress id CDATA #REQUIRED>


<!-- Domains Elements -->
<!ELEMENT Domains (Domain*)>
<!ELEMENT Domain (Geometry, Info)
<!ATTLIST Domain id CDATA #REQUIRED>


<!-- Geometry Elements -->
<!ELEMENT Geometry (Position, Dimension, Appearance)>
<!ELEMENT Position EMPTY>
<!ATTLIST Position xyz CDATA #REQUIRED>
<!ELEMENT Dimension EMPTY>
<!ATTLIST Dimension size CDATA #REQUIRED>
<!ELEMENT Appearance EMPTY>
<!ATTLIST Appearance colour CDATA #REQUIRED>

<!-- Info Elements -->
<!ELEMENT Info (Hits, Links?)>
<!ELEMENT Hits>

<!ATTLIST Hits uniqueHits CDATA>
<!ELEMENT Links (Link)>
<!ELEMENT Link EMPTY>
<!ATTLIST Link
    id CDATA #REQUIRED
    name CDATA #REQUIRED

>
```

The following is an excerpt from an example document that adheres to the DTD described above. The source document is an actual generated XML file from database information gathered during the development of the Cyberscape application. This map was generated for a user requesting a domain name at IP address 209.217.122.136. The full contents can be found in the provided file: "**CyberscapeMap_209.217.122.136.xml"**

*Excerpt:*

```
<Map  id="209.217">
 <User  currentposition="978.0 0.0 282.0" />
 <Street  id="209.217.122">
  <StreetSegment  id="209.217.122.136">
   <Geometry >
    <Position  xyz="978.0 -0.95 243.0" />
```

```
 <Dimension  size="4.0 0.1 3.0" sensorSize="4.0 5.1 3.0" />
 <Appearance  emissiveColor="1 1 1" diffuseColor="0.25 0.25 0.25" />
</Geometry>
<Info >
 <Hits  unique="13" />
</Info>
<InetAddresses >
 <InetAddress  id="209.217.122.136">
  <Geometry >
   <Position  xyz="981.0 0.0 243.0" />
   <Dimension  size="2 6 2" />
   <Appearance  emissiveColor="0.25 0.3 0.25" diffuseColor="0.25 0.25 0.25" />
  </Geometry>
  <Info >
   <Hits  unique="10127" />
  </Info>
  <Domains >
   <Domain  id="www.moses.cx">
    <Geometry >
     <Position  xyz="" />
     <Dimension  size="" />
     <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
    </Geometry>
    <Info >
     <Hits  unique="4870" />
    </Info>
   </Domain>
   <Domain  id="www.soulharvest.ca">
    <Geometry >
     <Position  xyz="" />
     <Dimension  size="" />
     <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
    </Geometry>
    <Info >
     <Hits  unique="143" />
    </Info>
   </Domain>
   <Domain  id="www.linuxgruven.ca">
    <Geometry >
     <Position  xyz="" />
     <Dimension  size="" />
     <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
    </Geometry>
    <Info >
     <Hits  unique="122" />
    </Info>
   </Domain>
   <Domain  id="www.kenhall.ca">
    <Geometry >
     <Position  xyz="" />
     <Dimension  size="" />
     <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
    </Geometry>
    <Info >
     <Hits  unique="78" />
    </Info>
   </Domain>
   <Domain  id="www.whymedia.ca">
    <Geometry >
     <Position  xyz="" />
     <Dimension  size="" />
     <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
    </Geometry>
    <Info >
     <Hits  unique="80" />
    </Info>
   </Domain>
   <Domain  id="webmail.linuxgruven.ca">
    <Geometry >
     <Position  xyz="" />
     <Dimension  size="" />
     <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
    </Geometry>
    <Info >
     <Hits  unique="4556" />
    </Info>
   </Domain>
   <Domain  id="ottawa-hs-209-217-122-136.s-ip.magma.ca">
    <Geometry >
     <Position  xyz="" />
     <Dimension  size="" />
     <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
    </Geometry>
    <Info >
     <Hits  unique="2" />
    </Info>
   </Domain>
   <Domain  id="cyberscape.whymedia.ca">
```

```
        <Geometry >
         <Position  xyz="" />
         <Dimension  size="" />
         <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
        </Geometry>
        <Info >
         <Hits  unique="153" />
        </Info>
       </Domain>
       <Domain  id="www.parkinsonphotography.com">
        <Geometry >
         <Position  xyz="" />
         <Dimension  size="" />
         <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
        </Geometry>
        <Info >
         <Hits  unique="88" />
        </Info>
       </Domain>
       <Domain  id="www.grantheckman.com">
        <Geometry >
         <Position  xyz="" />
         <Dimension  size="" />
         <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
        </Geometry>
        <Info >
         <Hits  unique="6" />
        </Info>
       </Domain>
       <Domain  id="www.alternity.ca">
        <Geometry >
        <Position  xyz="" />
        <Dimension  size="" />
         <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
        </Geometry>
        <Info >
         <Hits  unique="29" />
        </Info>
       </Domain>
      </Domains>
     </InetAddress>
     <InetAddress  id="209.217.122.137">
      <Geometry >
       <Position  xyz="975.0 0.0 243.0" />
       <Dimension  size="2 2 2" />
       <Appearance  emissiveColor="1 1 1" diffuseColor="0.25 0.25 0.25" />
      </Geometry>
      <Info >
       <Hits  unique="2" />
      </Info>
      <Domains >
       <Domain  id="hellmouth.mussar.com">
        <Geometry >
         <Position  xyz="" />
         <Dimension  size="" />
         <Appearance  emissiveColor="" diffuseColor="0.25 0.25 0.25" />
        </Geometry>
        <Info >
         <Hits  unique="2" />
        </Info>
       </Domain>
      </Domains>
     </InetAddress>
    </InetAddresses>
   </StreetSegment>
  </Street>
</Map>
```