# Atomsnet User Manual

Willem de Bruijn
Leiden University

July 2002

# Contents

# 1 User Manual

The atomsnet application has been designed to be easy to use for beginners, yet extensible enough to be interesting for more advanced computer users. In the introduction I will give an overview of the most basic functionality and how one can use the system 'out of the box'. More in-depth information on extending and personalizing the system is discussed in the other sections. That said, the introduction alone should suffice to help most people on the way.

Many different versions of the application have been released in the public domain during the development phase. Since many alterations have been made along the way I cannot describe every feature of every version. I will therefore restrict myself to the versions regarded stable. This manual therefore discusses solely atomsnet versions 1.0x, where the x can be any number.

## 1.1 Introduction

### 1.1.1 Prerequisites

Atomsnet relies on the Microsoft XML Parser version 3 or higher. This means that you need to have this package installed on your computer prior to running atomsnet. If atomsnet cannot open the XML parser it will show a message in the message window. Newer computers will have no problem, since the MSXML parser is part of Internet Explorer 6. If you do have an earlier version of internet explorer installed you should either upgrade the entire internet explorer package, or you can download the parser itself from microsoft.com

Atomsnet will only run on the Microsoft Windows operating system. It has been tested with Windows 98SE, Windows NT4 and Windows 2000 on Intel based computers.

### 1.1.2 Installation

Installing the application is very straightforward. There exist both a sourcecode download and a binary version. Application users should choose the latter package. This version of the application comes with a complete installer/uninstaller and can be installed without hassle within a few minutes.

After downloading the package, simply start the installer. During installation a few questions regarding package placement are asked after which the atomsnet application is installed on the user's computer. After this process a program link should have been added in the user's start menu.

In most cases no configuration is necessary to start working with the program. For safety the application will ask the user to check the settings upon starting the program for the first time, but it is generally safe to leave most settings as they are. The only setting that might be of interest is the 'remote server'. Atomsnet isn't preconfigured to connect to a network of active servers. If you do want to become part of an active p2p network you will have to add the IP address and portnumber of a running server here. A good source for server addresses is the discussion board at the atomsnet website.

Uninstalling the program can be accomplished by clicking on the start menu link below the application link. The automated uninstaller will then completely remove all references to atomsnet on the user's computer. Only when a user has added new files such atomsnet databases or webpages in the atomsnet directory will this location exist after uninstallation. Removing these files and deleting the directory has to be done by hand in that case. It is important, however, to run the uninstaller prior to deleting the directory by hand. Otherwise some references to the applications will remain on the local computer.

### 1.1.3 The Menus

Contrary to many Windows applications, the atomsnet user interface is split into separate menus. Upon starting the application the main menu is shown. The menu is shown in the next image.


Apart from the exit button, all of these buttons open a separate window.

The top botton, 'organize my data' allows a user to create an new database or open an existing one. When a database is opened files can simply be dragged from the windows explorer to the menu after which they will be added to the database. Be sure to save the database before you leave this menu. More detailed information on indexing data can be found in the 'indexing data' section.

The next button opens the network window. Here you can start and stop the network server. The menu has the same options for opening and closing database files as the indexing menu described above. This allows you to keep multiple databases for different uses. When closing this menu, the network is automatically shut down. Due to some network management done in the background it may take a while before the network is shut down, and therefore before the menu closes.

For debugging purposes all modules send human readable messages to the main applications. These can be viewed in a separate menu by selecting the 'show messages' button. In many cases this information will not be of interest to the user and therefore the button can be left unset. If you are experiencing problems or want to follow what's going on in the background this is an essential tool.

Finally, the settings dialog allows you to alter atomsnet's configuration. Be careful, however, in most cases these settings are correct and don't need to be altered.

### 1.1.4 Configuration Options

The configuration options can be set from the settings dialog. I will explain them in the order in which they are displayed.

First of all the location of the atomsnet homedirectory has to be given. This is initialized to the directory atomsnet was installed in. Changing this value can make atomsnet unusable, so please make sure that it is always set to the location of the application. Again, under normal circumstances this option can be left alone.

Second, the local IP address and portnumber can be chosen. Under normal circumstances the IP address is guessed by the application. It should only be altered if a user finds himself shielded from the rest of the network by a so called firewall. Configuring atomsnet for use behind a firewall is discussed in the 'sharing data' section below. The local portnumber is set to 80 by default. This is the normal webserver port and it is advised to use this port. In case another application is already running on port 80 this setting can be changed to the user's liking. Please note that atomsnet will show a message in the message window if it is unable to claim port 80.

The next option regards connecting to an active network of atomsnet servers. If you want to become part of a network you should add the address of a known server, together with its portnumber (usually 80) here.

Finally, you can set the maximum number of simultaneous connections. By default this is set to 5, but it can be any number from 0 to 25. The number of simultaneous connections has nothing to do with the users that access the computer from their webbrowser. It deals with the number of other atomsnet servers your server can contact directly. Since increasing this value means more background communication is allowed it can seriously degrade both your computer's responsiveness and your available network bandwidth. Setting this value to anything higher than 5 can be considered unwise in a highly active network. It is however useful in private networks that generate little activity, such as company intranets or in-home networks. Please note that setting this value to 0 disallows any connections to other servers, thus making your atomsnet application a stand-alone webserver.

## 1.2 Finding an item

To clients the network is displayed as a website. This means that although the atomsnet application can only run on Windows computers anyone can access the shared information. To search across a network you have to enter the address of a running server that is active within this network. This computer can have a hostname (e.g.: http://www.myserver.net) or an IP address (e.g. http://10.0.0.1). Webbrowsers guess that they should connect to port 80. This is why atomsnet servers are configured by default to listen on this port. In case the server has been reconfigured to listen at another port this has to be told to the webbrowser by appending the following:
:X

, where X is the portnumber. For instance, a server running on a computer with IP 192.167.12.14 and port 2000 can be accessed by typing http://192.167.12.14:2000 in the webbrowser's address bar.

When connected, you will see a webpage that can be either the standard atomsnet entrypage or a page created by the server's administrator. In the general case you will find both a search and multiple browse options. Searching the network works similar to other searchengines. Browsing is modeled to so called directories, such as Yahoo! (http://www.yahoo.com/) and the Open Directory Project (http://www.dmoz.org/ and http://directory.google.com/). Anyone accustomed to searching the web will be able to use the atomsnet. Because many computers have to be accessed for a request the waiting time can be longer than with regular searchengines, however. How long you have to wait before results are displayed is linked directly to the speed of the server, the size of the network and the bandwidth between the participating computers.

Although the program's functionality is extendible the basic browsing options are the following. Firstly, you can find files according to their type, this means all files in the network are grouped according to their nature. Searching by type (or more precisely MIME-type) uses a two level structure. You start by selecting the broad type, for instance video. After that a selection has to be made according to the precise filetype, for instance MPEG. It is, however, also possible to simply select video files and from there start a search, so that all video files regardless of their subtype that are of interest will be presented.

Another browse option is 'by category'. Categories are directly copied from the Open Directory Project format and display files not according to their type, but according to their content. This way it is for instance possible to find all files that concern football. Thousands of subcategories exits, making this a very finegrained browsetree. It is important to note, however, that this feature is experimental and not all files are categorized correctly. Hopefully the successrate will go up when new and improved plug-ins become available, but currently it is mainly useful for files where the filename contains the subject, or for mp3 audio (browse by artist).

The last browse option is 'by directory'. This option mimics the physical directory structure of the server you are accessing. It is added with the idea that people normally create their directory structure in a logical way. Using this option means relying solely on the archiving techniques of another person. It can be useful as an alternative to the browsing options discussed earlier, since many people index their files either by filetype (for instance a 'song' directory containing all mp3 files) or by category (for instance a directory containing all files belonging to a certain project).

When a file is selected, all characteristics can be shown by clicking on the filename. This way the mimetype location, the category and the directory location of the item are exposed, allowing you to browse similar files from this starting point. Furthermore, file size, date and additional information is made visible.

Since the network user interface is modeled closely to existing web resources I believe using the system is straightforward for most internet users. It is however possible for anyone that administers a server to alter the look and the working of his server to his personal liking. This means that certain features described above can be turned off, others can be added and the overall navigation can be rewritten completely. The usability of the network is completely in the hands of the server administrator. If you do not like a particular layout you can choose to 'hop' to another server until you find one that suits you. As a last resort, it is possible to create a personal entrypoint by installing the application on your local computer and connecting to the existing network.

## 1.3 Indexing data

Data can be indexed by accessing the 'organize my data' menu from the main menu. The first thing you need to do is open a file. This can be done in several ways. You can either open a previously created file or start with a blank file. When an older database is opened the system starts by checking the consistency of the database. This might take a while if you are loading a database containing many files.

When the database is loaded files can be added by simply selecting them in windows explorer or a similar program and dragging them to the menu. As soon as a file or selection of files (and/or directories) are dropped on the menu the applications starts to find information about the files through several options. The application relies on plug-ins, small separate programs, to find this information. The amount of information found and the time it takes to process depends largely on the plug-ins used. If you want to know what plug-ins the system uses open up the messages window before loading the database. Especially plug-ins that connect to other computers, for instance the google indexing plug-in that is part of the base package, will be active up to a few seconds per file. This makes the indexing process quite slow. It is therefore best to add up to a maximum of tens of files at once. Another reason to not overload the system is that the addition of data entries will take up quite a lot of memory space. As a reference, adding a thousand files can take up approximately 100 mega bytes of main memory. When all files have been added to the database make sure you save the file before exiting the menu, otherwise all new information will be lost.

## 1.4 Sharing data

Sharing the data can be accomplished by opening the 'share my data' menu. From that menu you can select to open a database. As soon as the file is successfully loaded the webserver is started. Only if the 'open file' buttons become grayed is the network active. Please open the message window and try again if the network is unable to load. This usually means that the local port is already occupied or the database cannot be loaded correctly.

As soon as the webserver is active a background process starts as well. This process tries to find other servers on the network and remains active as long as the webserver is active. Only when no remote server is given in the settings or when the maximum number of simultaneous connections is set to zero will this process stop immediately. How many connections are active can be seen at the bottom of the menu. Here a message can be found that usually reads 'no connections' upon network activation. During the active period other servers can be found, increasing the number of connections. This number will never exceed the maximum number of simultaneous connections as given in the settings dialog.

Leaving the network is accomplished by clicking the the 'close and return' button. It might take a while before the menu is actually closed. This is not a bug. During the waiting time the server is gracefully closing all active connections and closing the webserver. Since the system might be busy we have to wait a little while. Please be patient and wait for the system to return. Shutting down the application (for instance with the Windows Alt-F4 combination) will shutdown the network much faster, but can result in data inconsistencies.

### 1.4.1 Firewalls

If you do not know what a firewall is it is probably safe to skip this section. If you are aware that your computer is shielded from the rest of the active network, usually the internet, by a firewall some extra steps have to be taken to be able to run the atomsnet server.

There are two possible problem when a server is started from behind a firewall. Both problems can be dealt with and more often than not they exist simultaneously.

Firstly, it is possible that your computer does not have an internet wide IP address. This is made up for by having a firewall that acts as a masquerading server. Explaining how masquerading works is outside the scope of this document, but many good information sources can be found on the web. In short, masquerading makes your computer look like the masquerading server to everyone at the other side of the firewall. Since your IP address is unknown to people on the wide network they cannot access your computer. To circumvent this your server should make himself known by supplying the closest point on the wide network: the masquerading server. This is precisely where the 'local IP' option is for in the settings dialog.

Let me give a small example. Say your computer is behind a masquerading server. Your IP address is 10.0.0.2 and the server has address 100.0.0.1 on the wide network and 10.0.0.1 on the local network. While the addresses seem alike, 100.0.0.1 is a valid internet address, while everything in the range 10.X.X.X is reserved for local use. Anyone try to access 10.0.0.2 will therefore fail to reach you. If, however, you supply the internet address 100.0.0.1 others can find your masquerading server and all that needs to be done by this computer is to resend the data to your computer. This leads to the second problem.

Firewalls allow users behind the firewall to access computers outside their local network, but restrict strangers to access local computers. To allow users to access your atomsnet server you need to alter the firewall so that it allows others to access part of your computer, namely the atomsnet server. This can result in a serious security breach and has to be done with the utmost care. Ask permission from the firewall administrator before changing anything in the firewall! Since atomsnet listens on one port, namely the one configured in the settings window, you will only have to allow the firewall to forward data destined to this port to go through. Everything else should stay blocked.

In case of a masquerading firewall, remember that data is actually sent to the firewall, not to your computer. In this case the data should not only be allowed to travel through the server. You have to explicitly tell it that data destined for the firewall (in our example 100.0.0.1) on the same port as the one your computer is expecting it (for instance port 80) should be redirected to your computer. This means redirecting all requests going to 100.0.0.1:80 to 10.0.0.2:80 .

The portnumbers on the firewall and the atomsnet server have to be the same. This means that you should set your local port to a value acceptable for the firewall. Again, ask your firewall administrator for a suitable port, usually this will be a number greater than 2000.

Allowing external accesses to your computer can cause security breaches. Especially the configuration of the firewall is very important. Since atomsnet can only deal with security on its designated port all other open connections are the user's own responsibility. How atomsnet deals with network security is dealt with in the next section.

### 1.4.2 Security

For every program that has access to network resources security is a key concern. Atomsnet has been created as an easy to use program, meaning that configuration of the system is largely unnecessary. Therefore no human errors can make the system less secure (aside from the possibly needed firewall configuration discussed in the previous section).

The simplicity of the application is the strongest advantage of the atomsnet application. Since external users are only allowed basically four actions and by default denied others I believe the system to be very secure. All actions allowed use the HTTP GET protocol. Any other network requests are automatically declined. All requests result in read-only operations. I will now deal with safety regarding each allowed action.

**Accessing the database** Most user actions will concern retrieving information from the database. This information resides in main memory, so no disc access has to be given to the user. Only requests for data based on database item identifiers is accepted, therefore governing that no access is given to any information outside of the database. If you want to know what information is accessible please inspect a database by using an XML editor or text editor. In any case, no personal information is given.

Accessing the database should therefore not cause a direct security breach. Users should take notice that the following data is exposed: the directory layout as far as the files added to the database is concerned and the IP address(es) of the user's computer. This information can be considered of limited value to anyone trying to access your computer.

**Accessing webpages** Apart from querying the database, users can also access all files in the directory specified as the homedirectory in the settings window as well as all of its subdirectories. This is commonly the place where the atomsnet application and webfiles reside. The reason for this is that the atomsnet application has to be able to respond to requests for webpages, in the very least index.html . Since no access is given to locations using the 'parent directory' (..) locator, other parts of the computer are inaccessible to clients.

**Retrieving files** Shared files are made accessible as well, but these can reside anywhere on the user's harddrive. To make sure no other files can be viewed these shared files are only made available through their database identifiers. Since no direct filename based requests are allowed access is restricted to the files explicitly added to the database by the user.

**Finding network resources** The last action the server can be requested to undertake is to send a list of known servers. Replies to this request are simply lists of IP addresses. This should not lead to any security issues.

**Stability** The only possible security breach to my knowledge that can come from using the atomsnet application lies in the flooding approach. Flooding is generally used to get the computer in an unstable state. This can sometimes allow a user to get access to low level networking features *outside atomsnet*. Having an atomsnet server running does incur processor overhead. Misusing the system by sending a massive amount of simultaneous requests can degrade system performance to the point that the system becomes unresponsive and even unstable. In how far this can lead to security risks is dependent on the operating system. When viewed as such, any network server running on your computer can pose a threat, but a minor one at most.

### 1.4.3 Personalization

The atomsnet package comes with a standard built in webserver and some accompanying files. The webserver cannot be altered (in the binary distribution) for security reasons. You can change and replace the webfiles given with the main package to create your personal look and behaviour. Understanding of webtechnologies is necessary to change this data.

The standard package consists of four important files. These files can all be found in the installation directory of the application. Index.html is the file shown to a client when he accesses your server directly at http://www.myserver.net/ (exchange this hostname for your own hostname or address). This file basically explains how to use the system and provides links to the search and browse facilities. You can completely restyle your entrypoint by altering or exchanging this page with your own webpage. To alter the look of the main website you need to have a basic understanding of the Hypertext Markup Language (HTML). Information about HTML is publicly available on the web.

Apart from the index page, you can also alter the look and feel of the entire website, including the results the database send back by changing the file atomsnet.css . This file is written in the Cascading Style Sheets language. Please visit a website to learn more about CSS before altering this file. Basically, CSS tells the users webbrowser how a data item should be shown. For instance, the color, font and size of a text element can be chosen. By altering the data in atomsnet.css these and more layout features can be altered. For instance, changing
$background : white$
into
$background : black$
will make all pages have a back background. Of course, more interesting aspects can be changed.

If you want to change not how, but what information is displayed another file has to be changed. The file ane_html.xsl is is used to translate the raw database entry into a human readable format. This file is written in yet another language, XSLt. Again, please visit a website to learn more about this technology. In short, XSLt translates data from the raw XML form in which it is sent to the client into HTML document code. In our standard distribution, for instance, each browse result is shown to the user in a format resembling the Open Directory Project's style. You can display the data in a completely other format by simply changing this file.

The fourth file added to the distribution, corner.gif, is simply a picture containing the atomsnet logo. It is referenced from atomsnet.css . So, if you remove the reference to this image in atomsnet.css it is safe to remove the image.

Summarizing, the look and behaviour of the webserver is governed by standardized style documents. You are free to alter or exchange these files. Doing so does rely on understanding of the technologies used and can be a bit hard at first. Especially XSLt is a new technology and might pose a few problems.

## 1.5 Further information

I hope this manual has helped to understand the features of the atomsnet application. If you want to know more there are other resources publicly available. Apart from this document there exist a developer manual and a research document based on the application. Furthermore, the application is also being made available as sourcecode, meaning that you can view precisely how it works. For this an understanding of C++, Windows API's and webstandards is very useful.

If you have specific questions you can also ask them directly by placing a message in the discussion board at the atomsnet website. The address is http://atoms.sourceforge.net/ . Remarks and bug-reports can also be placed there.