A Beginner's Guide to URLs

What's a URL? A URL is a **Uniform Resource Locator**. Think of it as a networked extension of the standard *filename* concept: not only can you point to a file in a directory, but that file and that directory can exist on any machine on the network, can be served via any of several different methods, and might not even be something as simple as a file: URLs can also point to queries, documents stored deep within databases, the results of a *finger* or *archie* command, or whatever.

Since the URL concept really pretty simple ("if it's out there, we can point at it"), this beginner's guide is just a quick walk through some of the more common URL types and should allow you to be creating and understanding URLs in a variety of contexts very quickly.

File URLs

Suppose there is a document called "foobar.txt"; it sits on an anonymous ftp server called "ftp.yoyodyne.com" in directory "/pub/files". The URL for this file is then:

```
file://ftp.yoyodyne.com/pub/files/foobar.txt
```

The toplevel directory of this FTP server is simply:

```
file://ftp.yoyodyne.com/
```

The "pub" directory of this FTP server is then:

```
file://ftp.yoyodyne.com/pub
```

That's all there is to it.

Gopher URLs

Gopher URLs are a little more complicated than file URLs, since Gopher servers are a little tricker to deal with than FTP servers. To visit a particular gopher server (say, the gopher server on gopher.yoyodyne.com), use this URL:

```
gopher://gopher.yoyodyne.com/
```

Some gopher servers may reside on unusual network ports on their host machines. (The default gopher port number is 70.) If you know that the gopher server on the machine "gopher.banzai.edu" is on port 1234 instead of port 70, then the corresponding URL would be:

```
gopher://gopher.banzai.edu:1234/
```

News URLs

To point to a Usenet newsgroup (say, "rec.gardening"), the URL is simply:

```
news:rec.gardening
```

Currently, network clients like NCSA Mosaic don’t allow you to specify a news server like you would normally expect (e.g., `news://news.yoyodyne.com/rec.gardening`); this may be coming down the road but in the meantime you will have to specify your local news server via some other method. The most common method is to set the environment variable `NNTPSERVER` to the name of your news server before you start Mosaic.

HTTP URLs

HTTP stands for HyperText Transport Protocol. HTTP servers are commonly used for serving hypertext documents, as HTTP is an extremely low-overhead protocol that capitalizes on the fact that navigation information can be embedded in such documents directly and thus the protocol itself doesn’t have to support full navigation features like the FTP and Gopher protocols do.

A file called "foobar.html" on HTTP server "www.yoyodyne.com" in directory "/pub/files" corresponds to this URL:

```
http://www.yoyodyne.com/pub/files/foobar.html
```

The default HTTP network port is 80; if a HTTP server resides on a different network port (say, port 1234 on www.yoyodyne.com), then the URL becomes:

```
http://www.yoyodyne.com:1234/pub/files/foobar.html
```

Partial URLs

Once you are viewing a document located somewhere on the network (say, the document `http://www.yoyodyne.com/pub/afile.html`), you can use a *partial*, or *relative*, URL to point to another file in the same directory, on the same machine, being served by the same server software. For example, if another file exists in that same directory called "anotherfile.html", then `anotherfile.html` is a valid partial URL at that point.

This provides an easy way to build sets of hypertext documents. If a set of hypertext documents are sitting in a common directory, they can refer to one another (i.e., be hyperlinked) by just their filenames -- *however* a reader got to one of the documents, a jump can be made to any other document in the same directory by merely using the other document's filename as the partial URL at that point. The additional information (access method, hostname, port number, directory name, etc.) will be *assumed* based on the URL used to reach the first document.

Other URLs

Many other URLs are possible, but we've covered the most common ones you might have to construct by hand. At the top of each Mosaic document viewing window is a text field called "Document URL"; if you watch the contents of that as you navigate through information on the network, you'll get to observe how URLs are put together for many different types of information.

The current IETF URL spec is here; more information on URLs can be found here.

marca@ncsa.uiuc.edu