# United Nations Conference on Trade and Development

# Use of the Internet
# for Efficient International Trade

## Guide for SME Managers

### Chapter 1

### UNDERSTANDING THE
### INTERNET BEYOND THE BASICS

## Prepared by the UNCTAD secretariat

**United Nations**
**New York and Geneva, 2004**

# Chapter 1

# UNDERSTANDING THE INTERNET
# BEYOND THE BASICS

---

## 1. Objectives and organization of the chapter

The goal of this chapter is to explain and illustrate the major technical and commercial features of the Internet, which have potential impacts on businesses in developing countries.

This chapter is not an introduction to the Internet. It is assumed that readers have a basic understanding of the Internet and are reasonably proficient at using it.

This chapter lays the foundations upon which chapter 2 (*Introduction to e-business*) and chapter 3 (*Internet for international trade*) will build.

In this chapter we start by bringing to the fore *some key characteristics of the Internet* (section 3). The subsequent sections cover *services* which are *part* of the Internet; these services are the *web* (sections 4 to 7), *e-mail* (section 8), *file transfer protocol* (section 9), d*iscussion groups* (section 10), *IP telephony* (section 11) and r*eal time communication tools* (section 12). Last but not least, *Internet-based searches* ( section 13) deals with several of the above-mentioned services of the Internet.

## 2. Introduction

**The Internet is a network of several hundred million computers which are interconnected and located all over our planet. It is as simple as that.** Technically speaking, the Internet is not an impressive innovation. What is spectacular is the fact that, for the first time in history, there is a truly global and very dense computer network.

**Within the Internet, there is an array of services that are interlinked**. The most popular of these services are the *worldwide web* (*www* or *web* for short) and *electronic mail* (or *e-mail* for short). In other words, e-mail and the web (as well as the other services of the Internet) are *parts* of the Internet. In addition to e-mail and the web, there are many other Internet-based services such as Internet telephony (also known as Voice-Over-IP or IP telephony), file transfer protocol (FTP), discussion groups, chat, instant messengers (IM), mailing list server and Internet fax. All these services are covered in this chapter.

**Despite its enormous size, the Internet continues to grow**: in 2000, there were 388 million Internet users (6.5 per cent penetration rate); this number grew to 623 million (10.2 per cent penetration rate) in 2002.[1] The penetration rate is defined as the number of Internet users divided by the total population. The numbers of Internet users are necessarily rough estimates. Indeed, there is no precise and commonly accepted definition of who can be considered an Internet user. The fact that many people have access to the Internet but that they never or almost never use it is one typical issue. Another major cause for imprecision is that sources of data are often privately owned Internet Service Providers that tend to inflate their own figures for commercial reasons.

**Internet penetration rates strongly vary from country to country,** as shown by the following figures for 2002:[1]

- Sweden: 57%
- United States: 55%
- Brazil: 8%
- Bhutan: 1.4%
- Senegal: 1.0%
- Cambodia: 0.2%

# 3. Some key characteristics of the Internet

In the list of characteristics below, we start with general characteristics and gradually move to more specific ones. These characteristics apply to the Internet in general; however, for the sake of simplicity, we tend to illustrate them using examples of websites.

**Key characteristic 1: The Internet is a truly worldwide medium**. This means that any computer connected to the Internet can have access to all (public) Internet services such as websites of the world. For example, an Internet user in Cambodia, the Lao People's Democratic Republic or Bhutan can have access to valuable information about a trade fair in Japan such as the *Northeast Asia Business Fair "NAB-Messe"* through its website *http://nab-messe.erina.or.jp* (see picture).



---

[1] International Telecommunication Union (ITU). Internet indicators: Hosts, Users and Number of PCs 2000 and 2002, 15 December 2003. See www.itu.int/ITU-D/ict/statistics/.

**Key characteristic 2: The Internet is fully delocalized**. In other words, its different players can be located anywhere in the world. Let's take the example of a website with the following seven types of players.

*Player type 1: The website publisher* is the formal owner of a website. Let's suppose he is located in Bhutan.

*Player type 2: Website builders (or designers)* create websites, technically and graphically. Website publishers sometimes build their website themselves, but let's suppose that our Bhutan-based publisher has appointed a Thai company to build his website. Communications between the publisher and the Thai website builder can be by e-mail (even though face-to-face contacts could have been easier). The Bhutanese publisher may have selected a Thai company given its particular expertise in building good websites. To have access to some skilled low-cost workforce, the Thai company may subcontract some portions of the work (e.g. database programming) to two companies: one in Myanmar and the other one in India.

*Player type 3: Registrars*. Each website has an address (e.g. BhutanExample.com) called a domain name. This Internet address must be registered by an official registrar (see section 5 on domain names management later in this chapter). Let's suppose that the Thai website builder registers the above domain name (on behalf of the Bhutanese publisher) with *GANDI Sàrl* (*www.Gandi.net*), a French registrar known for offering very low prices.

*Player type 4: Website hosts* are the companies where many websites are *physically* hosted (see section 7 in this chapter). Website hosting is highly specialized work requiring high-speed dedicated lines, permanent maintenance and qualified staff. In website hosting, there are important economies of scale and US-based hosts tend to have the most competitive offers (e.g. $100/year for hosting simple sites). Therefore, the Bhutan publisher and his Thai service provider jointly decide to use an US-based host. For several technical reasons (back-up, load-balancing etc.) it is not uncommon for important sites that hosting is done simultaneously in several locations.

*Player type 5: Website content creators/updaters.* To create the content of a website is often a major effort requiring significant management attention. Once the content has been created, it normally requires regular updates. If the website is well built, the *technical operation* of loading the content in a website (as opposed to the *editorial creation*) is easy to do, even by unskilled clerks. Let's suppose that our website is about *authentic Buddhism* and that our Bhutanese publisher has hired a team of 25 university professors living in 12 different countries to create the content of this website. Furthermore, the site is built in such a way that each professor has his own section; naturally, each professor can load content directly into his section.

*Player type 6: Website technical updaters* make technical modifications to a site, as required by website publishers. To make minor modifications is normally easier than to build a site from scratch. Let's suppose that our Bhutanese publisher has hired a freelance woman in Bhutan to do this sporadic job; however, this freelance woman left for the Netherlands, where she is studying for a Master's Degree. Since she already invested some time to understand how the website has been built, the Bhutanese publisher decided to continue working with her.

*Player type 7:* Finally, once the website is up and running, *website viewers (or surfers)* can naturally be located in <u>any country</u>.

Finally, please note that the term "webmaster" is not clearly defined: it can refer to the *website content creators/updater, the website builder* or *the website publisher.*

**Key characteristic 3: The Internet is a free-of-charge and open technology.** In addition, free tutorials can be easily obtained from the Internet itself. The "source codes" of websites (i.e. how websites have been programmed) are to a large extent viewable by any surfer. Thus, the Internet-related know-how is not the preserve of a few powerful developed countries but is fully accessible by all countries, including least developed countries (LDCs).

How to see the "source code" of a web page (for instance of *www.OnlineTies.com* from *OnlineTies.com Sàrl*)? Using *Microsoft Internet Explorer*, select *View* then *Source*. The resulting "source code" is shown below (see below).



**Key characteristic 4: The Internet has no central ownership and no central management.** However, some people (e.g. within the framework of the *World Wide Web Consortium, www.w3c.org*) issue recommendations regarding the technical definition of the Internet. It must also be said that some large companies such as *Microsoft* do not always fully adhere to these recommendations and sometimes define their own rules.

**Key characteristic 5: To learn how to use the Internet is relatively easy.** *As a user*, to learn the basics of the Internet (i.e. e-mail, worldwide web) requires from one hour to one day; to master all the Internet services takes two to three months.

However, to become an Internet professional, much more time is required. For instance, only one to ten days are required to learn how to build simple websites but at least one year must be devoted (for people with IT background) to learning how to make sophisticated websites. It should be noted that highly sophisticated websites are created by *teams* of specialists.

**Key characteristic 6: Access to the Internet is relatively inexpensive and access prices tend to drop rapidly.** As an example, in 2003, prices of public access (i.e. access from cybercafe; see picture) in Phnom Penh (Cambodia) hovered around $0.50 per hour.



*Internet Café in Phnom Penh, Cambodia*

These access prices remain certainly unaffordable for many people in LDCs; however, these prices are much lower than those of comparable services such as traditional international phone calls. In this regard, the costs of Internet-based international phone calls (about $0.06/min in an LDC)[2] are much lower than the costs of traditional international calls (about $2/min in an LDC)[3]. Furthermore, the quality of these Internet-based phone calls from LDCs was already quite acceptable in 2003. Such Internet calls can be made using one's own computer or from a cybercafe, provided it is legally acceptable. To make Internet-based phone calls, specific user-friendly applications such as the one of *Net2Phone, Inc.* (*www.Net2Phone.com*; see opposite) can be used.



In the above example, phone calls are made from a computer to any regular phone (i.e. PC-to-Phone). However, making PC-to-PC Internet-based phone calls is even much cheaper because only Internet access costs need to be paid. To do so, tools such as Microsoft Net Meeting can be used (see specific section in this chapter).

---

[2] Net2Phone.com Inc. Rates. 16 December 2003. See http://dcs.net2phone.com/products/commcenter/.
[3] Source: field study in Cambodia and Bhutan in 2003 as well as in Nepal, Myanmar, Madagascar and Togo in 2001.

**Key characteristic 7: The Internet can accommodate multimedia features.** For instance, websites, e-mails or FTP sites can have:

- Texts (not only in English but in many other languages)
- Images
- Audio (see example below taken from *GoCambodia.com*, *www.GoCambodia.com*: clicking on the links will trigger the download of corresponding Cambodian songs)



- Video (see example below taken from the Japanese TV channel *Japan Broadcasting Corporation NHK*: it is possible to view videos of the news from the website *www.nhk.or.jp*)

**Key characteristic 8: The Internet is interactive**. Internet users not only receive information but can also provide their own feedback. By doing so, users can for instance participate in opinion polls, make specific keyword-based searches or register in trade fairs. A real example of Internet interactivity is the website of Cambodia's translation company *Pyramid* (*www.pyramid-e.com*; see below) which gets online inquiries from potential clients all over the world.



Finally, in an increasing number of countries, citizens can now vote and fill their tax declarations online. This generates substantial monetary and time savings for both Governments and citizens. The example below is the website of the French fiscal authorities (*http://www.impots.gouv.fr*). To have such an online service, a prerequisite is to have a system of *digital signatures* (see below in this chapter) to authenticate citizens.



**Key characteristic 9: The Internet is "always on" but it is not 100 per cent reliable.** For instance, websites can be momentarily unavailable and e-mails can get lost. The latter is due to a combination of technical and human reasons.[4]

**Key characteristic 10: There are very powerful tools which allow surfers to make searches within the Internet.** This topic will be covered in depth in section 13 of this

---

[4] According to empirical evidence gathered by the author, about 1 to 2 per cent of e-mails get lost.

chapter but, in a few words, Internet searches span over a very vast amount of data, have pertinent ways to rank search, and are fast and mostly free.

**Key characteristic 11: Some information about Internet users can be collected on the spot through the Internet itself.** Taking the example of the web, what kind of data about web surfers can be collected by website publishers?  For example:

- The *approximate location* (i.e. more or less in which city/region) of surfers
- The *type of browser* (incl. the language version) used by surfers
- The *webpages visited by surfers in the past* (including the transactions they have carried out)

Importantly, a surfer's identity (i.e. his real name) cannot be known unless this identity has been correctly declared at some point by the surfer *himself*. However, as we will see under characteristic 12, surfers can also surf anonymously provided they resort to very specific techniques and tools.

Technical precisions: when visiting websites, surfers leave "raw" traces on web servers. Per se, these raw traces are difficult to use. However, there is an array of statistical packages which help make sense of these raw traces. For instance, *Webtrends* (*www.netiq.com/webtrends*), a well-known statistical package for websites, provides the following outputs:

- *Most and least requested pages;*
- *Average user session length,* i.e. long users stay on average on a website;
- *Top entry and exit pages,* i.e. from which pages surfers access and leave a website;
- *Single access pages,* i.e. on which pages of a website visitors access and exit without viewing any other page;
- *Top paths through sites,* i.e. the paths surfers most often follow when visiting a website;
- *Most downloaded files,* such as a piece of software or a music clip;
- *Information about the most frequent visitors,* i.e. organizations/companies, types of domain names, countries, regions within a country, etc.
- *Top authenticated users,* i.e. who is visiting the website most often (applies only when a website requires a user name/password);
- *Traffic summary*: day by day, by day of the week or by hour of the day;
- *Total hits, failed hits, cached[5] hits* and (server's or surfer's) *errors*;
- *Used bandwidth;* the bandwidth is the amount of data that can be sent/received by a transmission line, measured in bits per second;
- *Top referring sites and search engines,* i.e. where do surfers come from? (In other words: do they type the website address? Do they come through an advertising banner? Do they come from search engines or other links?);
- *Top keywords entered in search tools* to find the web site;
- *Browsers and operating systems* most used by surfers;

---

[5] Technical note: The *cache* is the temporary storage area for frequently accessed or recently accessed data. Having certain data stored in cache speeds up computer operation but distorts raw access data about website access.

- *Visiting spiders;* a spider (also called *web crawler*) is software that visits websites and indexes the pages present in those sites. Search engines use spiders to build up their databases. Analysing spiders' traces allows website publishers to know by which search engines their site has been (re)indexed.

Information about web surfers has many of commercial uses. For instance, website publishers can monitor precisely and in real time the impact of an online promotional campaign (using the "top referring site" feature, as explained above). They can also make targeted promotional offers to specific surfers.

To understand about traces left by surfers, we recommend visiting (the English version of) the *Commission Nationale de l'Informatique et des Libertés (CNIL)* outstanding interactive website (*www.cnil.fr*; see picture). *CNIL* is a French independent administrative authority; computer-related technological watch is one of its key missions.
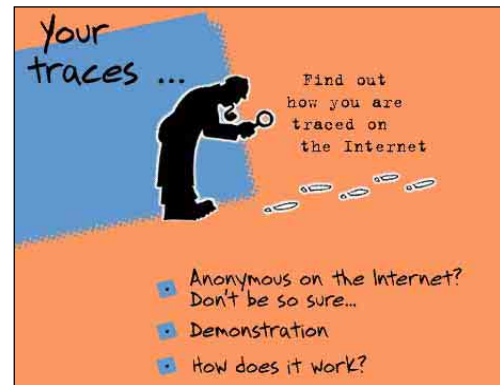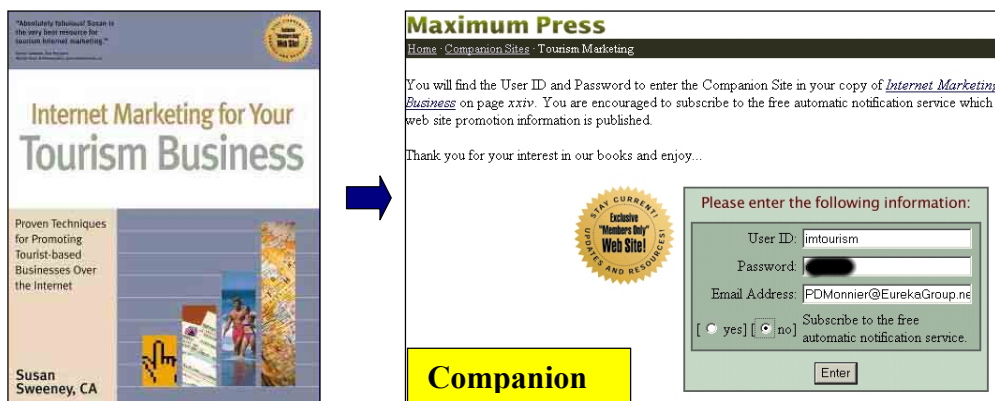
**Key characteristic 12: The Internet can be anonymous in many regards**. For instance, web surfers, e-mail senders and author of websites can remain anonymous. To do so, the use of specific software from, for instance, *Anonymizer, Inc.* (*www.Anonymizer.com*) and *Steganos GmbH* (*www.Steganos.com*) are instrumental. However, to ensure full anonymity, advanced specific expertise is required.

**Key characteristic 13: Information can be made available on the Internet nearly instantaneously** while, in the case of paper-based publication, much more time (and money) is needed to print and distribute information. As an example, the website of UK-based newspapers, for example *The Financial Times* (*www.FT.com*; see below), is updated many times within a single day; besides, articles about hot topics have several online versions since these articles are updated/augmented as soon as new information/comments become available.

**Key characteristic 14: Access to Internet services can be customized.** As an example, a website can be:

- *Accessible to all*: this is the "normal" (i.e. open, standard, public) *Internet*
- *Accessible to "insiders" only*: in this case, we have an *Intranet, i.e. a type of closed network.*
  Example: A Ministry of Trade and Industry Intranet would be accessible only to (some of its) officers
- *Accessible only to " insiders" and " close associates":* in this case, we have an *Extranet i.e. a type of broad but closed network.* Example: A Ministry of Trade and Industry Extranet would be accessible to (some of its) officers as well as to registered exporters.

Another example of an Extranet is the "companion website" *(www.maxpress.com/tourism/entry.html*; see below) of the book entitled *Internet Marketing for Your Tourism Business* (see below), published by *Maximum Press*. The role of a "companion website" is to *support* a book, a newspaper, a TV programme, a sporting event, etc. In order to have full access to the companion website below, it is necessary to enter a username and a password that can only be found inside the book.



**Key characteristic 15: Once on the Internet, information cannot be fully censored.** Governments can, however, make access to some websites difficult and punishable. This is a highly complex topic. As an example, a Government can ask all the i*nternet a*ccess providers (IAP) in its country to block a (probably long and constantly changing) list of censored sites; in addition, a Government can also ask all IAP in its jurisdiction to collaborate in tracing surfers visiting prohibited websites.

**Key characteristic 16: The Internet works on many technical platforms.** This is actually a small technological revolution. As an example, the Internet can be used with different operating systems such as Windows, Apple Macintosh OS and open-source Linux (see respective logos below).

**Key characteristic 17: The major costs of defining and implementing an Internet business strategy are not the technical costs** (software, hardware, programming, hosting). These costs are often relatively minor, except for very sophisticated websites. The main costs incurred in defining and implementing a web business strategy are as follows:

- Management reflection
- Employee training
- Website *content* creation/update
- Website promotion
- *Business process re-engineering (BPR)*[6] definition and implementation.

Depending on the types of websites, total investments and expenses in building a website vary from next to nothing to several million dollars.

**Key characteristic 18: The key competencies required in order to define and implement an Internet business strategy are not technical competencies** but a combination of the following know-how:

- A good understanding of the company/organization and its environment as well as a good sense of business/marketing
- Familiarity with the Internet, in particular as a commercial tool. In this regard, it is essential to understand how target users react to the Internet.
- Technical know-how, including knowledge of graphics/layout applied to the web
- Editorial competence

These competencies are rarely found in a single person or even in a single company. Consequently, outsourcing is often required.

# 4. A few introductory words about the web

Briefly defined, **the *worldwide web* (or *web* for short) is one of the services on the Internet**. It is a collection of hyperlinked documents (or *web pages*) formatted mostly in *HTML* (see later in this chapter) and made (usually) publicly available on computers (or *web servers*) around the world. The information on these servers can be viewed or accessed with a browser such as *Microsoft Internet Explorer* or *Netscape Navigator*. The web started in 1993 in Geneva (Switzerland) and its inception is commonly credited to *Tim Berners-Lee*. We shall now see two characteristics which are specific to the web rather than the Internet as a whole.

**It can be relatively easy to build a website, but it is usually difficult and/or expensive to make it known.** To become known, some websites such as the Internet-based diversified retailer *Amazon.com Inc*. (*www.Amazon.com*) spend more than one hundred million dollars a year in promotional expenses.[7]

---

[6] *Business process re-engineering (BPR)* can be defined as the fundamental redefinition of business processes leading to radical improvement in organizational effectiveness. New tools such as the Internet often allow spectacular BPR.

[7] More precisely: in 2002, marketing expenses (net of cooperative reimbursements) were to the tune of US$ 125 million (but US$ 180 million in 2000). Source: Amazon.com Inc. 2002 Final Annual Report. See http://media.corporate-ir.net/media_files/irol/97/97664/reports/2002_Annual_Report_FINAL.pdf.

To promote their websites, well-known established companies have an edge: they can for instance take advantage of their strong brands, existing advertisement campaigns and existing infrastructure (e.g. retail outlets).

However, as we will see in chapter 2, many SMEs both in developed and in developing countries found creative ways to promote their website at low costs.

**A website can be personalized, i.e. its content and presentation can be user-specific.** There are many business reasons why a website publisher may wish to build a personalized website. Some examples:

(a) A multilingual website can automatically load pages in the language of surfers;

(b) A website can have personalized advertisements, sales suggestions or special offers in order to increase its commercial effectiveness;

(c) A website can remember some specific data (i.e. personal data, delivery address, billing address, bank account number, specific wishes) of its registered surfers so that re-ordering can be executed expediently.

Technically, web pages can be personalized on the basis of:

- *Data entered by surfers themselves,* e.g. during registrations, purchases, comments, etc. These data are called *explicit* data.
- *Traces left by surfers in web servers.* Traces can indicate approximate location, language of browsers and operating systems, etc. These data are called *implicit* data.

In the example below, we can see an advertising banner in French (see inside the red ellipse) on the website of the English-language newspaper *Financial Times* (*www.FT.com*). What happened is that the website identified that the surfer is located in a French-speaking city of Switzerland; consequently, a French-language Swiss advertising banner is being displayed.
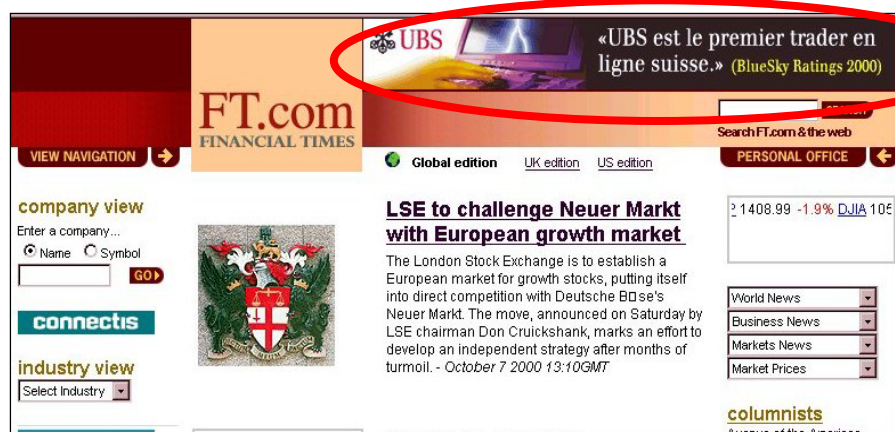


Another example of website personalization is the *MyYahoo* (*http://my.yahoo.com*), i.e. the personalized part of Yahoo! Inc. (*www.yahoo.com*). Surfers can easily personalize several aspects of *MyYahoo* (e.g. colour, content and layout) using the control panel (see picture).
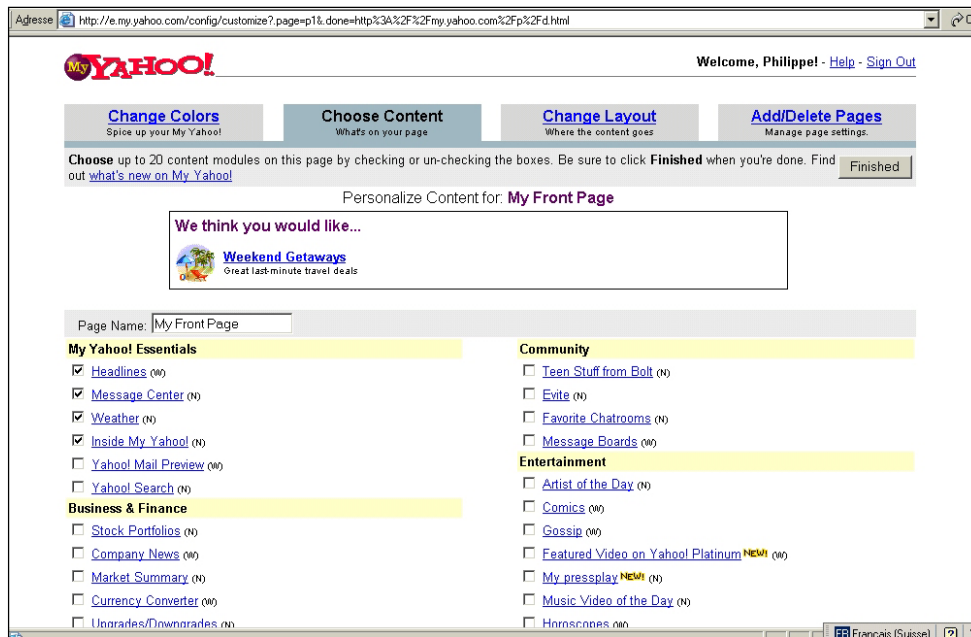
# 5. Domain names management

**What is an Internet domain name?**

A typical *website address* (also known as *Internet address* or *Uniform Resource Locator or URL*) looks like this: ***http://www.any_name.com***. Let's consider each element:

- *http* means "this is a website";
- *www* can be replaced by anything and can even be suppressed;
- *any_name.com* is the **domain name.**
- *.com* is the *Top Level Domain name (TLD),* also known as *(domain name) extension.*

Some TLD are *generic TLD* (*gTLDs* for short) such as .com, .net, .org, .biz, .info, .name and .aero. This means that they do not refer to any particular country. *.com* is the most common domain name extension; it was originally meant for <u>com</u>merce-oriented websites, but any person or organization can now reserve .com domain names.

Other TLD are called country code TLD (ccTLD). They are specific to a given country. Some examples are:

- .co.jp; .go.jp; or.jp; .ne.jp;  … (Japan)
- .com.ph; .gov.ph;  .org.ph;  … (Philippines)
- .com.bt; … (Kingdom of Bhutan)
- .com.kh; … (Kingdom of Cambodia)
- .de (Germany)
- .ch (Switzerland)

**Some conventions about website addresses:**

(a) If a web address looks like Amazon.com, iic.edu.kh, Eureka.jp or Wadeda.ac.jp, it is safe to assume that the full web address is http://www.Amazon.com, http://www.iic.edu.kh, http://www.eureka.jp or http://www.Waseda.ac.jp. Should the typical www not be used, it is necessary, for the sake of clarity, to write a more complete address such as labs.google.com or even http://labs.google.com.

(b) Domain names are not case-sensitive. However, for better readability, it is common to use capital letters at the beginning of each word. For example, *GoCambodia.com* (or *www.GoCambodia.com*) is easier to read than *gocambodia.com* (or *www.gocambodia.com*).

**Why is it essential to properly register one's domain names?**

We will illustrate this concept with the website of *Kuensel*, the sole Bhutanese newspaper. In 1997, *Kuensel* launched a website whose address was *www.kuensel.com* (see below).
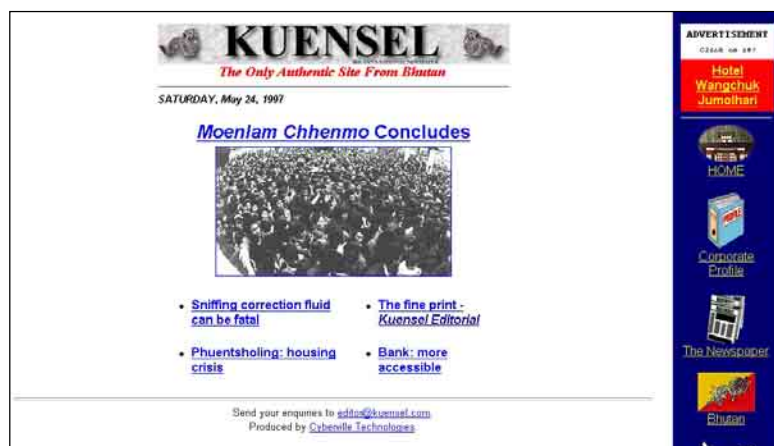


However, in 2003, surfers who typed www.kuensel.com ended up in a … commercial pornographic website. What has happened? Why did this happen? Since the Bhutan-based designer of this website did not own a credit card to make the payment needed to register the Kuensel.com domain name, he had to ask an agent to do it on his behalf (see below; some data have been blacked out).
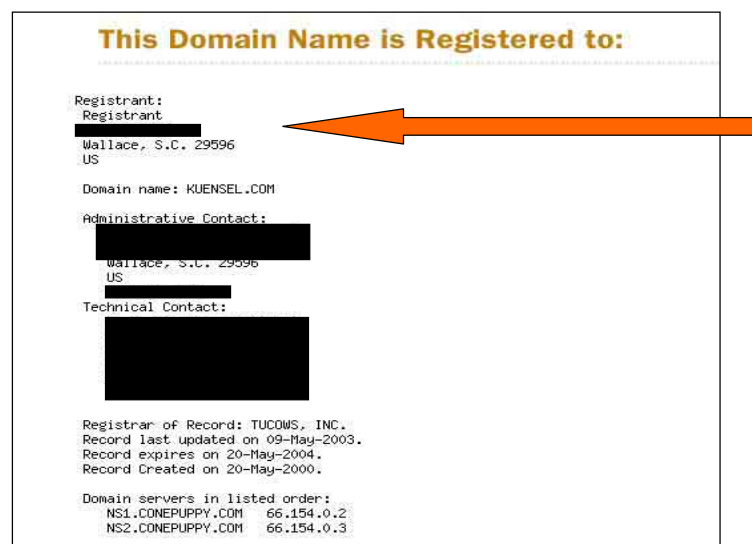
Generally speaking, to register a domain name through an agent is a disadvantage because such an agent has at any time the possibility to redirect surfers to any other website. It is as if the Coca-Cola trademark was not owned by Coca-Cola Corporation but by an agent who could use this trademark at any point in time and prevent Coca-Cola Corporation from using it. Furthermore, the agent can also forget to renew the domain name registration upon its expiry so that this domain name can be registered by anybody else on a "first come first served" basis (this is reportedly what happened in the case of Kuensel.com).
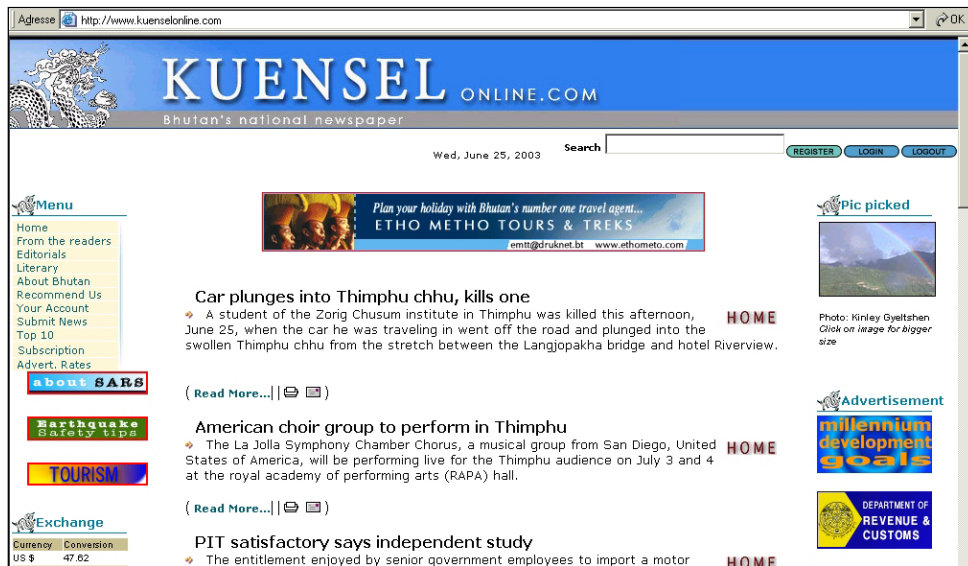


Now *Kuensel* has a new website with a new domain name *www.kuenselonline.com* (see above).  For the same reason as the first time, this new domain name *kuenselonline.com* was again registered through an agent (see below; some data have been blacked out), which may again present similar risks. Being aware of these risks, Kuensel plans to renew the registration under Kuensel's own name from 2004 onwards.
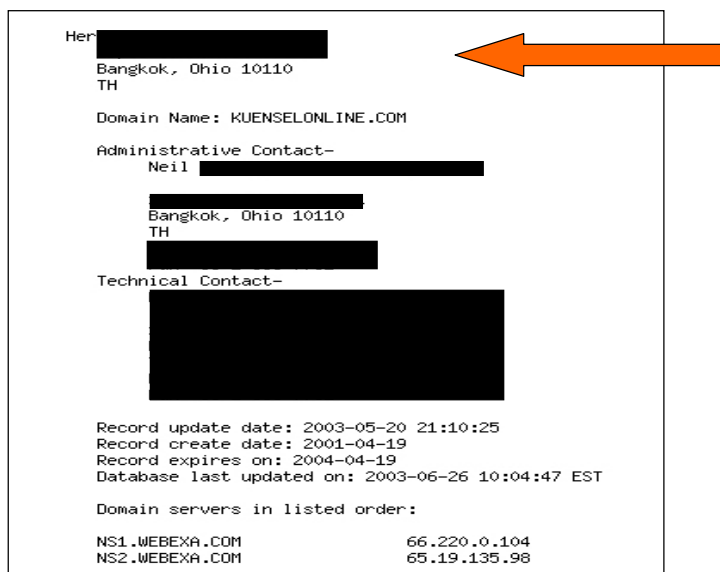


In this section, we will see how SMEs in both developed and developing countries could prevent this type of problem.

**Where to register domain names?**

gTLD domain names can be registered online with *any* ICANN-accredited registrar. *ICANN* stands for *Internet Corporation for Assigned Names and Numbers*. There are about 150 *ICANN*-accredited registrars and they are located in all continents. The complete list can be found on the following web page: *www.icann.org/registrars/accredited-list.html* (see below).
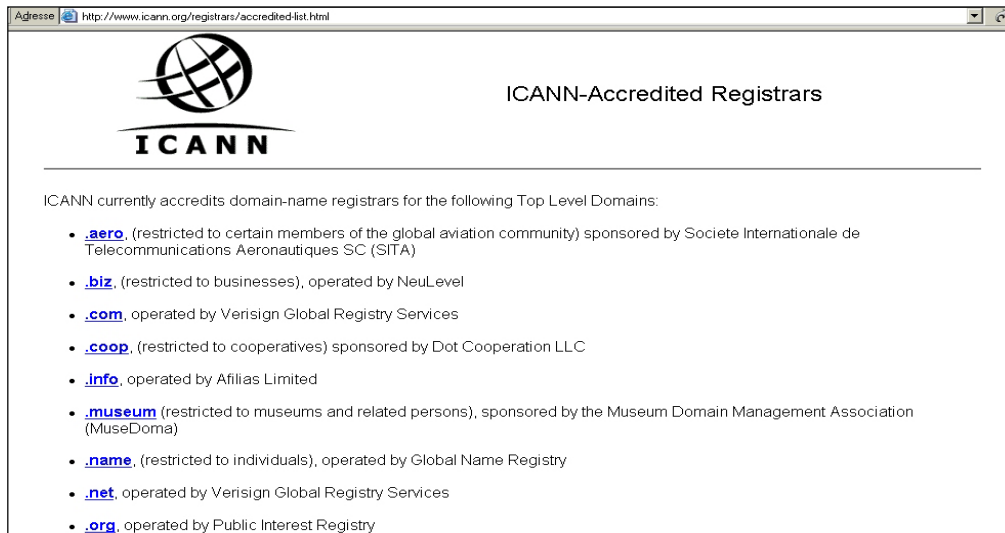


ccTLD domain names must be registered with national registrars. There is normally only one national registrar per country.

Usually, domain name reservations follow the "first come first served" rule unless there is a clear case of brand name usurpation.
To register a domain name usually costs from $10 to $100 per year (as of 2003),[8] depending on the registrar; ccTLD domain names tend to be more expensive, probably because of the absence of competition among ccTLD registrars.

**How to find out by whom a given (gTLD) domain name has been registered?**

By making a search using a so-called *WHOIS* server. Websites of registrars normally have links to these types of servers. Examples include the webpage *www.internic.net/whois.html* from the *Internet's Network Information Center (InterNIC)* or the webpage *www.register.com* from *Register.com, Inc*. From these WHOIS servers, it is possible not only to find information about the organizations/persons who have registered given domain names but also to verify whether a domain name is still available.

Information about country code domain names can be found from the website of *UWHOIS, Inc*. (*www.uwhois.com*, see picture).

---

[8] Based on the author's experience.
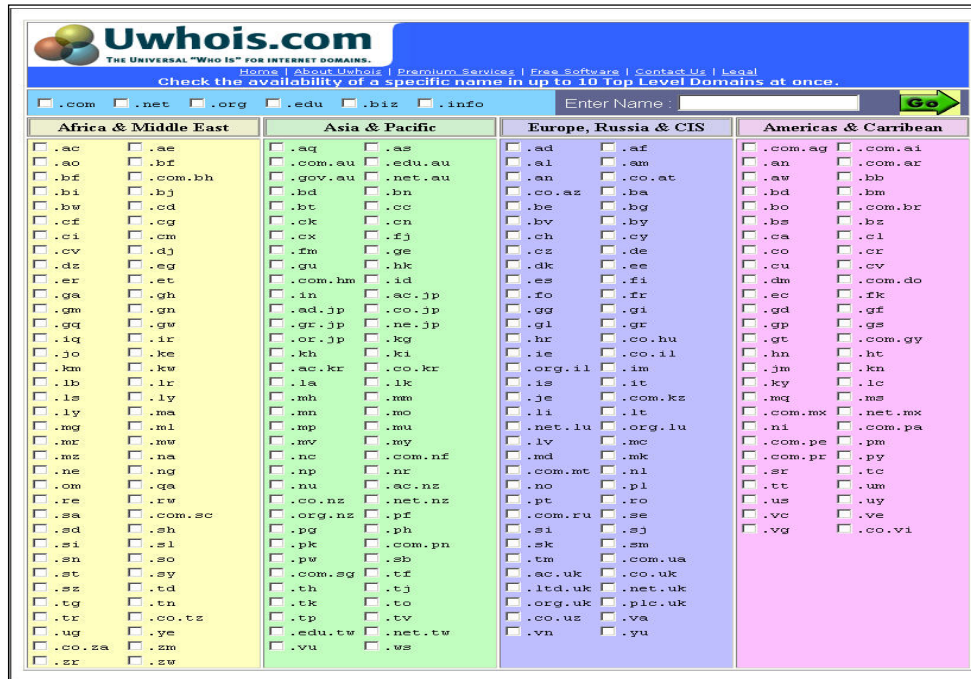
**Is it better to reserve a country code domain name (e.g. SAMPLE.com.kh) or a generic domain name (e.g. SAMPLE.com)?**

Whenever possible, both (and more) should be reserved, as will be explained later in this section. The real question is therefore *which domain name should be your main one* (i.e. the one to be communicated/promoted). Actually, both country code domain name and generic domain name have their own advantages.

Advantages of *Country Code Top Level Domain Names (ccTLDs):*

- More exotic
- More domain names are available
- Local payment method can be used (no need for credit cards or other international payment methods)

Advantages of *Generic Top Level Domain Names* (gTLDs):

- Usually cheaper (from $10 year)
- Shorter extension (e.g. .com rather than .com.kh)
- More prestigious

**Which practical rules/tips should be observed when choosing a domain name?**

**(a) Domain names must be easy for the target audience to pronounce, understand, remember and spell.** Usually, the shorter the better: SAMPLE.com is better than SAMPLECompany.com.  Furthermore, it may be a good idea to use English names if international markets are targeted.

**(b) People who do not know one's domain name should be able to guess it**. In other words, if one's company name is SAMPLE INC, it make sense to reserve (at least) SAMPLE.com, provided it is available.

If an organization only has a *sub-website* (i.e. a website which is a part of another website), it typically has a very long website address. For instance, the *Cambodian Handicraft Association* (see below) has the following URL: *www.bigpond.com.kh/users/wthanchashop* because its website is a subwebsite of *www.BigPond.com.kh* owned by *Cogetel Limited*. However, this association could reserve a domain name such as *cyberCHA.org* and have the address *www.cyberCHA.org* rerouted to their current *www.bigpond.com.kh/users/ wthanchashop*. Therefore, surfers would just need to remember and type *www.cyberCHA.org*. CHA.org or CHA.com would have been even better (i.e. shorter), but these domain names were already taken as at mid-2003.
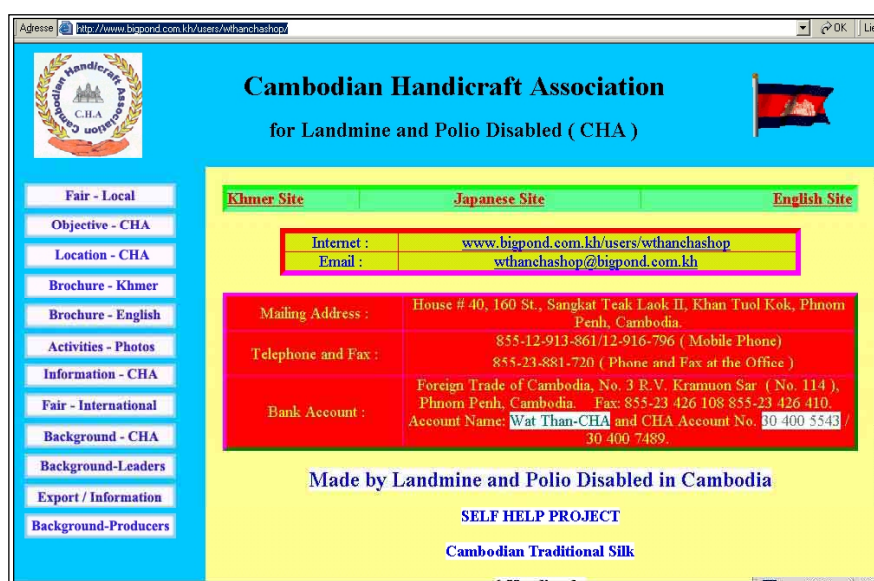


**(c) Reserve several similar domain names and have them rerouted to one single website.** As an example, the company *Peek & Cloppenburg* has reserved the following French domain names: *p-c.fr, peek-cloppenburg.fr; peekandcloppenburg.fr, peekcloppenburg.fr, and peeketcloppenburg.fr*; obviously, they have also reserved generic domain names and domain names from other countries. If this rule is followed, even if one's surfers misspell one's domain name, they will still end up in the proper site. Companies or organizations, which did not take these precautionary measures, end up having bad surprises. For instance, *www.WhiteHouse.gov* is the official site of the American President's Office; however, if surfers mistakenly type *www.WhiteHouse.com* (published by WhiteHouse.com), they end up in a pornographic website; moreover, if they mistakenly type *www.WhiteHouse.org* (published by Chickenhead Productions, Inc),* they end up in a parody website.

**(d) Use the same domain name for both website address and e-mail addresses**. As an example, if the domain name of an organization is *unctad.org*, all members of this organization should have e-mail addresses looking like *name@unctad.org.*

**(e) For promotional reasons, always include the extension when writing one's domain name.** If the extension is not mentioned, surfers may not find one's website easily. A good example is *Amazon.com*'s logo (see picture).

**amazon**.com.

**(f) Be sure to be the official registrant**. Many companies in developing countries ask agents to reserve domain names on their behalf; unfortunately, these agents often reserve domain names under their own name and, once traffic has been built, reroute these domain names to other sites.

Why are agents so often used in developing countries to reserve domain names? Because to reserve domain names typically requires credit card payments and most SMEs in developing countries do not have credit cards. Two things can be done to solve this problem:

- Make sure that agents reserve domain names *under their clients' names*.
- Reserve domain names in registrars who accept payment methods such as cheques, wire transfer/TT or demand draft (DD); usually, Indian registrars, e.g. *Polar Software Limited* (*www.SignDomains.com*) and *Direct Information Pvt. Ltd.* (*www.directi.com*), accept these payment methods.

**(g) Monitor timely renewal dates**. Domain names are reserved for a given number of years with unlimited possibilities of renewal. However, if not renewed at the right time domain names become available for anybody to reserve/use. Registrars usually (*but not always*) send reminders before renewals are due but do not rely exclusively on them.

# 6. Website building management

The **objective** of this section is not to explain in detail how to build websites because this topic is highly specialized and technical. In addition, many good books and Internet resources cover this topic. The objective of this section is to make sure that readers are able to choose the type of website they need given their budget and objectives; this section should also help readers to dialogue with website builders and to control their work.

**Some golden rules to build effective websites** are as follows:

- *Communicate, do not decorate*: it is more important for a website to convey clear messages to its target audience than to be a piece of art.
- *Consider technical outsourcing* because to build good websites usually requires talents rarely available internally. Besides, it is important not to build everything from scratch but to use what already exists. Technical resources (e.g. software program) for site builders can be found in websites such as *SiteInspector.com* (*www.SiteInspector.com*).
- *Make one's website easily updatable* so that new content can be loaded very easily even by unqualified clerks.
- *Make one's website easily scalable* so that its size can increase without causing major technical redesigns.
- *Make one's website (nearly) permanently operational* because unavailable websites (e.g. under construction, under reconstruction or inaccessible because of an overloaded or malfunctioning server) make a bad impression.

- *Consider outsourcing of general content*. To create and update good-quality content is expensive. Many companies plan very rich content but, owing to lack of resources to update it, they end up having outdated content. Consequently, instead of creating general content oneself, it usually makes more business sense to create links from one's website to specific webpages of generic content providers such as online newspapers and press agencies.
- *Instil confidence,* especially if one's company is not at all well known and located in a developing country. See some suggestions below.

**To instil confidence through a website is not easy but is important** because a minimum level of confidence is required to generate business transactions. To create confidence is especially critical for websites of SMEs based in developing countries because, by nature, these companies are less trusted than well-known large corporations based in developed countries. Measures to instil confidence include:

- Professional presentation (including no spelling or grammatical mistakes)
- Picture/background of management/staff (no anonymity). A good example can be taken from the website of *ThaiGem.com* (*www.ThaiGem.com*, see below)



- Addresses/pictures of offices, retail outlets
- Association with trusted/known organizations. Example: logo display of professional associations, large clients (provided they agree)
- Customer testimonials (ideal: eBay.com Feedback Summary; see below)
- Professional-looking (i.e. short) domain names
- Link to the commercial registry where one's company is registered.

**How to build websites, technically speaking?**

As mentioned earlier, **the heart of a website is its source codes. Websites can have different technical levels**. Grossly simplified, we can categorize websites into five levels (from the simplest to the most complicated one). These levels are not mutually exclusive.

*Level 1: Static sites*. These sites also called – *brochureware* – provide simple information as if they were paper brochures.

*Level 2: Interactive/transactional sites*. These sites offer the possibility of carrying out interactive operations such as taking orders, collecting opinions or making online payments.

*Level 3: Database-based sites.* The webpages of these sites are not created in advance but are generated on the fly using data stored in databases. These sites are suitable for coping with large amounts of (dynamic) data. A typical example is the website of the *Swiss Federal Railways* (*www.rail.ch*; see opposite): when a surfer makes a timetable query, the results are taken from databases and not from an existing static page.

*Level 4: Personalized sites (also known as one-to-one sites).* These sites have content/outlook which are surfer-dependent (refer to "Key characteristic 2" above).

*Level 5: Data-mining-based sites.* These sites extrapolate surfers' tastes and provide customized content based on these extrapolations. (*Data mining* can be defined as the process of using statistical techniques to discover subtle relationships between data and the construction of predictive models based on them). Regarding websites, an example could be as follows: if a data-mining-based site *anticipates* that a surfer is likely to appreciate hiking, special promotions about trips in Nepal will be proposed to him.

**There are several ways to build websites.** These methods are usually combined.

**Method 1: To write manually HTLM and other codes.** This method provides full flexibility but is rather difficult and time-consuming.

**Method 2: To use an (HTML) editor to produce HTLM and other codes.** This method provides full flexibility, allows time-saving but is rather difficult. This is usually the preferred way used by professional web builders.
An example of an HTLM editor is *WebCreator* (a tool written by *Kevin Lynn Brown;* see for instance *http://www.siteinspector.com/shops/webmaster/htmledit.shtml*), which can be found below. By using the keys at the top, HTML codes can be created very fast.

**Method 3: To use a WYSIWYG[9] editor.** This method is adequate for simple, less professional websites. It is very easy and provides relatively good flexibility. However, web builders do not have complete control of the codes; therefore, results looks less professional.

Here is an example of a website created with *Microsoft FrontPage*, a popular WYSIWYG editor. *Microsoft FrontPage* has the appearance of a word processor, but it actually generates HTML codes. The example below is taken from *OnlineTies.com Sàrl* (*www.OnlineTies.com*).
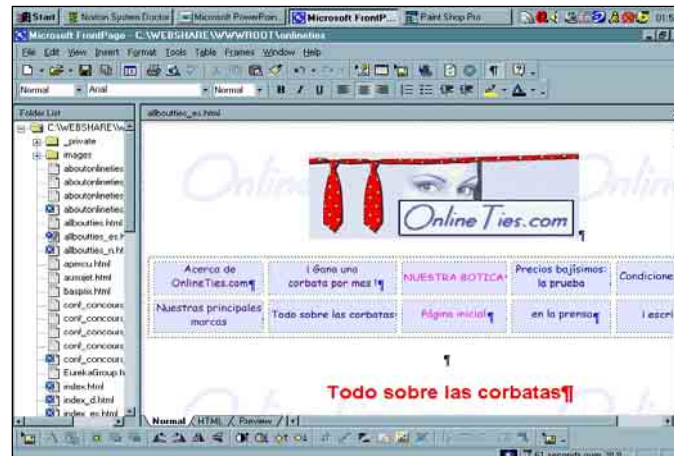


**Method 4: To use special software** such as Macromedia's Flash to enhance the graphical appearance.

**Method 5: To use template-based packages** (see e.g. *www.Geocities.com* from *Yahoo! Inc.*, *http://store.yahoo.com* from *Yahoo! Inc.*, *www.stores.ebay.com* from *eBay Inc.* or *www.mercantec.com from Mercantec, Inc.*). With these packages, it is very easy to build websites, including commercial-transaction-enabled websites. Flexibility is, however, quite limited.

An example of one of these template-based packages is **eBay Stores** (www.stores.ebay.com).



eBay Stores is part of *eBay* (www.eBay.com), the leading auction site (see chapters 2 and 3). eBay Stores are linked to eBay auctions and/or direct sales. As of 2003, tenancy fees were US$ 9.95/month; in addition, a percentage of transaction-based commissions has to be paid. Payment mechanism is *PayPal.com* (www.PayPal.com; see chapter 2) with payments accepted from about 50 countries. Escrow services (see chapter 2 as well) are recommended for transactions over $500. As an option, "ID Verify" allows eBay merchants to establish their proof of identity.[10]

---

[9] *WYSIWYG* is an acronym for *What You See is What You Get*. To be realistic, *we* should rather use the term *What You See Is More Or Less What You Get.*

[10] How does ID Verify work? eBay members' (i.e. merchants' or buyers') personal information is cross-checked against consumer and business databases for consistency. Concretely, eBay members are asked to enter information such as name and date of birth, and to identify certain instalment and credit accounts, and their

*Uthipul* (see below, www.ebaystore.com/uthipul) is an example of a Cambodian website which has been created using eBay Stores.



**Method 6: To use (highly) sophisticated e-commerce packages** such as *osCommerce*[11] (*www.osCommerce.com*), a free-of-charge product. These types of packages offer great flexibility (because they can be modified/configured at will) but normally require time and good computer skills.



The above example comes from *OnlineTies.com Sàrl* (*www.OnlineTies.com*) and was built using oSCommerce. Interestingly, this website used to be built with a combination of *Microsoft FrontPage* and *Mercantec;* however, to improve the design appearance and cope with larger transaction volumes, the website has been fully rebuilt using a combination of an *HTML Editor* and *osCommerce.*

---

associated monthly payments. When the information has been successfully verified, eBay members receive an ID Verify icon.

[11] Technical note: osCommerce i*s* an *open source* software working with (open source) *Apache, MySQL* and *PHP.*

# 7. Web hosting management

Every website must be physically hosted on a *web server*, which is a (powerful) computer permanently connected to the Internet through a *dedicated (high speed) line*. Web servers need permanent technical attention. Fortunately, website publishers do not need to own and run their own server. Actually, most website publishers *rent space in servers* of specialized companies known as *web hosts* or *web presence providers*.

Website publishers have three broad hosting options:

(a) To rent server space from a web host (this is termed **hosting** or "managed hosting" in the event of sophisticated hosting). This is the preferred option for most cases.

(b) To own one's own server(s) but to have it/them on the premises of the web host (this is termed **housing** or **collocation**).

(c) To have **one's own server**(s) and dedicated line on one's own premises. This is a high-end expensive solution very much appreciated by technical people; however, business-wise, this solution can only be justified in rare cases of major websites.

To compare web hosting offers, many parameters need to be understood. Typical parameters are mentioned below (this webpage is taken from *Bluewin*, *www.hostcenter.com*; by clicking on a parameter, surfers will be able to see a corresponding definition).



Web hosts typically provide a *control panel* to website publishers. A typical control panel looks like the one of Bluewin's HostCenter *(www.HostCenter.com):*

**Importantly, a website can be hosted anywhere in the world**, irrespective of the location of the website publisher, website builder or registrar. For an SME in developing countries, a key question is: should we host our website locally or in a foreign country such as the United States?

**Local hosting** has some advantages:

- Local payments (no need for a credit card or other international payment methods)
- Perhaps more influence on the service level if the web host is located close to the website publisher
- "Political correctness" (i.e. support to local companies).

However, top-notch **large international hosts** tend to offer some distinct advantages, mostly thanks to their critical mass:

- Higher reliability
- Higher security (e.g. against *hackers,* i.e. malicious intruders)
- Better services available 24 hours a day, 365 days a year
- Broader range of services
- Higher access speed (at least for international surfers)
- Competitive prices

# 8. Understanding e-mails beyond the basics

**Electronic mail (or e-mail for short) is the most used service of the Internet, along with the worldwide web.** According to IDC, the number of e-mails sent in 2002 is 31 billion per

day and this number is expected to reach 60 billion a day by 2006.[12] E-mails are most useful for business, especially in developing countries; indeed e-mails allow fast, practical and low-cost (about one US cent per e-mail) telecommunications.

There are **two basic ways to use Internet-based e-mails:**

**(a) Using a regular e-mail (client) software** such as *Microsoft Outlook, Microsoft Outlook Express or Pegasus Mail*

**(b) Using a web browser-based e-mail service** such as the well-known *Yahoo Mail* (http://mail.yahoo.com) or *Hotmail* (www.hotmail.com). There are actually many less popular or more local browser-based e-mail services such as Bhutan's *DrukNet Mail* (http://mail1.druknet.bt) or Cambodia's *Everyday Webmail* (www.everyday.com.kh; see picture).

**Regular e-mail (client) software has specific advantages** such as:

- Offline work is easier. (This is especially important if slow and/or unreliable (dial-up) Internet connections are used).
- Better functionalities (e.g. possible use of advanced functions such as spell checker of the Microsoft Office Suite).

However, **web-based e-mails also have their own advantages**:

- No separate e-mail client software is needed
- Very handy use for travellers, for instance using Internet cafes or laptop computers.

E-mails offer significant advantages, but several elements must be taken into account to fully use this tool.

To ensure e-mail **confidentiality** (i.e. to make sure that e-mails cannot be read by unauthorized people during their transmission), it is necessary to **encrypt** messages. The way to encrypt e-mails is described in a subsequent paragraph.

To ensure that senders and receivers are properly identified **(authentication)**, it is necessary to use **digital signatures.** Authentication is essential given that it is very easy to create e-mail accounts with fake names and addresses. Naturally, important e-mails such as contracts require the proper authentication of contracting parties. Digital signatures are issued by reputable bodies (called *Certification Authority)* such as government entities or well-known private companies (i.e. *VeriSign*, *www.verisign.com*; see picture).

---

[12] International Data Corporation (IDC) (2002). Worldwide E-Mail Usage Forecast, 2002-2006: Know What's Coming Your Way.

Digital signatures can be bought (approximate cost: US$10-20 per year) and they also provide the possibility of **encrypting** messages and ensuring their **integrity** (i.e. to make sure that messages are not *tampered* with during their transmission).

Digital signatures have the same **legal validity** as handwritten signatures in many countries. However, the real challenge – especially for SMEs in developing countries – is not only to be in compliance with laws but also to have enough funds to sue a dishonest (foreign or domestic) counterpart and/or to ensure that court decisions are (internationally) enforced.

**The transmission speed** of e-mails is usually but not always very fast. Delivery speed is slower (and can even take more than one hour) if networks are very busy and if messages are large, for instance because of bulky attachments.

The delivery of e-mail is **not fully guaranteed**. About one to two per cent of e-mails get lost. This loss rate increases in the case of heavy traffic and large messages. To prevent this problem, senders can ask receivers to confirm the receipt of a message by returning it. (Automatic confirmations of e-mails are not reliable because they depend on the technical capabilities of the receivers' e-mail software as well as on the receivers' willingness to allow their e-mail software to send automatic receipts).

To prevent a receiver **from repudiating an e-mail** (i.e. claiming that he did not receive it) requires special (complicated) systems of "registered e-mail". More practical methods include:

- Requesting an e-mail by way of confirmation (i.e. to ask that a receiver confirm receipt of an e-mail by returning it)
- In extreme cases, having the police scrutinize the receiver's hard disk.

**E-mails as such cannot contain viruses or worms** per se, but the following attachments of e-mails can:

- Microsoft Word, Excel, Power Point, macros
- Any executable files (i.e. files whose extension is .exe)
- Any audio and video files

Safe files include files ending with .txt, .pdf, .jpg or .gif. However, there are many complex tricks to make a harmful file look like a safe file.

It is very easy to buy millions of e-mail addresses (obtained without the consent of their owners) and to send **massive amounts of unsolicited e-mails.** For this reason, about 50 per cent of e-mails are now unsolicited mass e-mails known as **spams**. To fight against spam, it is possible to use software such as *MailWasher* (*www.mailwasher.net*) or to take legal action in

countries where spams (or at least some forms of spam) are illegal. But the best way to prevent spams is to avoid posting one's e-mail address on the Internet (websites, newsgroups). This is, however, in contradiction with the goal of being easily reachable. A compromise could be to have only one generic contact e-mail address per website.

In order to **mass e-mail,** several methods can be used:

**(a) Regular e-mail software (e.g. Microsoft Outlook) with distribution lists** (see example below)



It is, however, usually better to insert the distribution list in the BCC[13] field (rather than in the TO field) so that each recipient does not see who all the other recipients are.

**(b) Specific offline tools:** There are many software programs specialized in mass e-mailing. They can be easily found by making a search on the Internet. It is also possible to combine *Microsoft Word* with *Microsoft Outlook*. Details of how to do this depend on the specific versions of these software, but extensive help is provided by the help function of *Microsoft Word.*

**(c) Specific online tools:** *(Mailing) List Server* is an ideal solution for large mass e-mailing operations. A *list server* (or a *mailing list server* to use the complete name) is a server on the Internet (with its corresponding software) that processes requests to subscribe and unsubscribe to mailing lists and handles the distribution of e-mail to everyone on the mailing list. A list of existing list servers (and their associated mailing lists) can be found at the following Internet addresses: www.Topica.com of *Topica Inc.)* and www.tile.net/lists of *SparkNET Corporation*. For those interested in running their own list server, the following websites can be consulted:

- *Microsoft's* www.bcentral.com (click on *Sales & Marketing* and then on *E-Mail Marketing* and on *List Builder*)
- *L-Soft international, Inc.'s* www.lsoft.com (product name: *ListServ*)
- *Topica Inc's* www.topica.com (product name: Email Publisher)
- *Great Circle Associates, Inc.'s* www.greatcircle.com/majordomo (product name: *Majordomo*).

---

[13] Abbreviation of *blind carbon copy,* which means that the original addressee(s) will not know that a copy of a message has been sent to the people in the BCC field.

## 9. Transferring large amounts of information using FTP

*File Transfer Protocol (FTP)* is the ideal Internet service to transfer large files over the Internet. Files are uploaded to and downloaded from so-called *FTP servers* (i.e. specific powerful computers permanently connected to the Internet).

Business-wise, FTP allows companies to develop new commercial services. For instance, in chapter 2, we will see the example of *Digital Divide Data*, a fast-growing Cambodian company specialized in providing accurate data entry services to foreign clients. To transfer files from and to their clients, *Digital Divide Data* uses primarily FTP.  Another example: many companies load all their technical manuals in FTP servers. These manuals can be either accessible *anonymously* (i.e. to all without having to use specific user name and password) or to authorized users only (e.g. to clients or employees). A list of *anonymous ftp sites* can be found at www.tile.net/ftp from *SparkNET Corporation*. Typical anonymous FTP site (addresses) look like ftp://ftp.cisco.com (*Cisco Corporation* in the United States) or ftp://ftp.cfi.waseda.ac.jp (*Waseda University* in Japan).

Files can also be transferred as attachments of e-mails. However, FTP offers many more possibilities such as virtually unlimited file sizes and permanent availability of files. Moreover, interrupted downloads/uploads can resume where they have been stopped; this is very handy for countries where connections are slow and/or unreliable.

To transfer files using FTP, a software called **FTP client software** is needed. This software is very easy to use. *CuteFTP* from *GlobalSCAPE* (www.GlobalScape.com; see how it looks in the screen capture below) and *Ws_ftp* from *Ipswitch Inc* (www.ipswitch.com) are among the most used FTP client software.
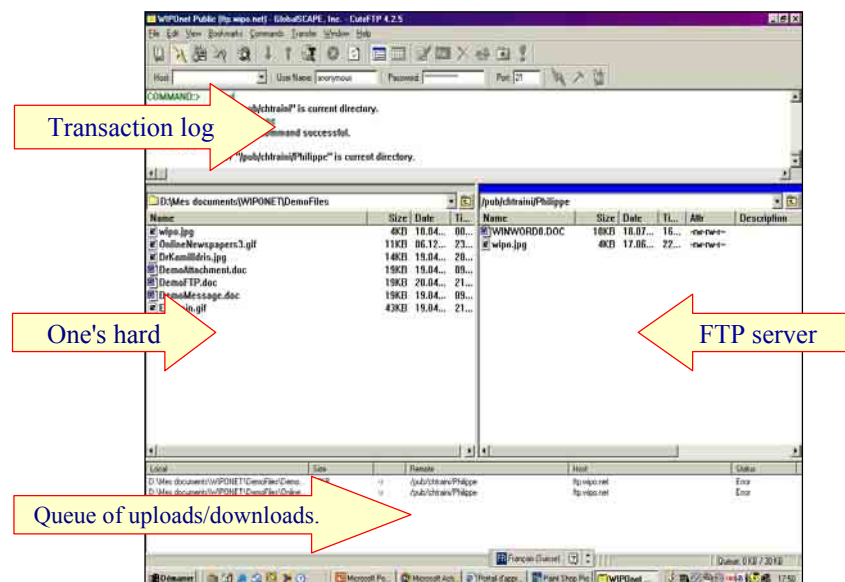


Actually, it is also possible to download files from a FTP server using a regular browser in such a way that most users would not even notice that files are downloaded from a FTP server rather than from a Web server.

It costs approximately US $20-100 per year to rent space (about 100 Mb) in a FTP server. Basically all *web hosts* provide this service.

# 10. Sharing ideas and finding out
# what consumers think using discussion groups

**Discussion groups** (also known as *newsgroups* or just *groups*) are bulletin boards on which Internet users can read and post messages. Millions of people participate in probably over 100,000 discussion groups.[14] There are discussion groups about virtually any subject.
Discussion groups are classified into a number of categories such as:

- *alt*. : Any conceivable topic.
- news. : News...
- *biz*. : Business products, services, reviews…
- *sci*. : Applied science, social science...
- *humanities* : Fine art, literature, philosophy...
- *comp*.: Hardware, software, consumer info...
- *rec*. : Games, hobbies, sports...
- *soc*. : Social issues, culture...
- *talk*. : Current issues and debates...
- *misc*. : Employment, health, and much more...

**Complete newsgroup addresses** look like this: <u>*rec.climbing*</u> (i.e. a newsgroup about climbing as a leisure activity) or <u>*soc.culture.laos*</u> (i.e. a newsgroup about the culture of the Lao People's Democratic Republic).

**Discussion groups** can be:

- **Moderated** (so that messages are screened by a moderator before making them accessible to all) **or not**;
- **Public** (i.e. everybody can have access) or **private**;
- **Permanent** or **temporary** (i.e. related to a short-lived event such as elections)

**Business-wise,** discussion groups are excellent ways to:

- *Share ideas* (this is very much used by technical people)
- *Find out what people really think about certain topics.* This is an especially powerful way to do it since newsgroup users tend to be very frank given their relative anonymity (because many people use false identities). Two practical examples: *Microsoft* can find out what people may dislike about their products; tour operators in Bhutan may discover what people really think about trekking in Bhutan (see screen capture in the section covering *Google groups* later in this chapter).

One of the easiest way to **access newsgroups** is to use "Google groups" (see section "Exploiting Internet search tools" later in this chapter).

---

[14] Precise statistics in this area are lacking, however.

# 11. Slashing telecommunications costs using IP telephony

**IP telephony is an Internet-based telephony.** IP stands for *Internet protocol* (a *protocol* is a *computer language*). IP telephony is also known as **Internet Telephony** or **Voice over IP (VoIP)**.

The beauty of IP telephony is that it requires an investment five to eight times less than the classical *PSTN* (Public Switch Telephone Network) technology, according to the *International Telecommunication Union (ITU)*.[15]

Practically, IP telephony allows users (primarily in developing countries) to make **very cheap international phone calls**, provided it is legal. Price savings are particularly significant in countries where monopolies keep prices of regular phone calls high.

As an example, IP telephony service provider *Net2Phone* (www.net2phone.com; see screen captures below) charged the following rates as of December 2003:[16]

- From Cambodia to Japan: US$ 0.059/min
- From Cambodia to the United States: US$ 0.049/min
- From Cambodia to Switzerland: US$ 0.08/min
- From Cambodia to Bhutan: US$ 0.86/min.

Regular Internet connection fees (say US$ 1 per hour) must be added to the above fees.
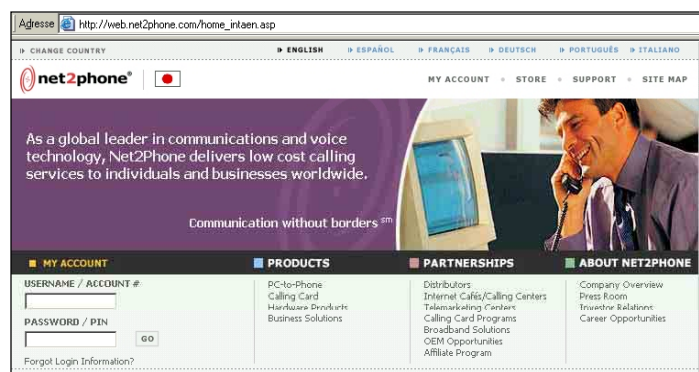


These fees have to be compared with regular telephony fees of several US dollars per minute to make international phone calls from many LDCs.

Using IP telephony, the **quality of the communication** ranges from barely acceptable to very good. The voice quality is basically affected by:

- Hardware components
- Software components
- Backbone network (i.e. speed of data transfer)
- Network traffic
- Other technical elements.

---

[15] International Telecommunication Union (ITU) (2000). Thailand: IP Telephony and the Internet, p. 6 (available online www.itu.int/osg/sec/spu/ni/iptel/countries/thailand/thailand-iptel.pdf).
[16] Net2Phone.com Inc. Rates. 16 December 2003. See http://dcs.net2phone.com/products/commcenter/.

To sum up, equipments with above-average capabilities are required in order to have good-quality communications.

An easy way to take advantage of IP telephony is to use this service in (properly equipped) cybercafes. Many of them in developing countries and in LDCs (e.g. in Phnom Penh) offer excellent quality of communication. An alternative is to use one's own multimedia-equipped PC, an Internet connection and a specific software downloadable from websites of IP telephony service providers such as US-based *Net2Phone* (www.Net2Phone.com) or India-based *Net4India* (www.phonewalared.com; see below).

Some payments must be made to these IP telephony companies. Since they also target developing countries, they typically accept an array of payment methods other than credit cards, such as wire transfers/TT, demand draft (DD), money orders and cheques. Imported *calling cards* can also be purchased locally from some cybercafes, e.g. in Bhutan and in Cambodia.



The solutions which we have mentioned so far are from a PC to a regular telephone. These solutions are called **PC-to-Phone**. But there are even much cheaper ways to have (international) Internet-based voice communications. These solutions require that both parties use simultaneously a computer connected to the Internet, and they are therefore termed **PC-to-PC** solutions; we will cover them in the next section, i.e. *Instant messengers, chats* and *online collaboration.* Price-wise, users of these PC-to-PC solutions only pay Internet connection costs (for instance, about US$ 0.5 to US$ 2 per hour in some Asian LDCs in 2003).[17]

Finally, it is also possible to use the Internet to **send and receive faxes** (see for instance a site such as www.eFax.com from *j2 Global Communications).* It is, however, much more economical to scan documents, edit them if needed and send them as an attachment of e-mails or using FTP.

# 12. Instant messengers, chats and online collaboration tools for real-time advanced communications

*Instant messengers, chats* and o*nline collaboration* are different types of Internet-based PC-to-PC tools designed for **communication in real time**. To some extent, these tools are a sort of simplified *videoconference* (i.e. a meeting between people at different locations relying on video technology as the primary communication link).

Originally, these three types of tools were rather different, but there is a growing convergence among them, both in terms of functionalities and user groups. For instance, all of them can now have voice and video functions.

---

[17] Source: field study in Cambodia and Bhutan in 2003 as well as in Nepal in 2001.

**Instant messengers (IM)** have the particularity that users can know which of their colleagues, business partners or friends using the same system are available online. If they are available, they can be instantly contacted.
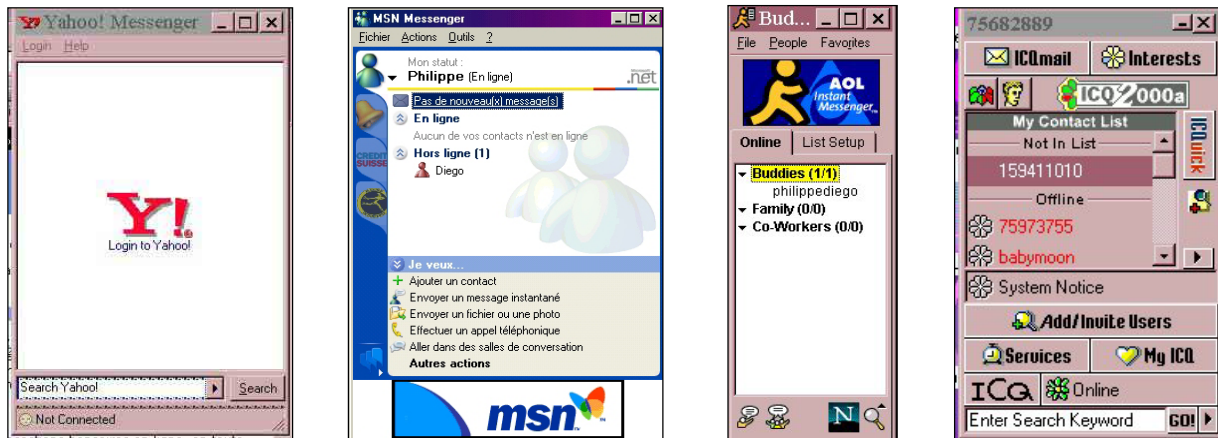


There are four main IM software (see above to see the way they look): *Yahoo Messenger (http://messenger.yahoo.com), Microsoft Messenger (http://messenger.msn.com), AOL Netscape Instant Messenger (www.aim.com)* and *ICQ (www.icq.com).* They are incompatible themselves. All of them have a free downloadable version.

**Chats** used to be mostly a game for teenagers but now they are used more broadly. Some chats can be used with a simple browser (examples includes Bhutan's *DrukNet Chat* www.druknet.bt (see below) and *SwissOnline AG*'s www.SwissTalk.ch), while many of them require downloading previously a specific software (the most popular one is *IRC* from *mIRC Co. Ltd.*: www.mirc.com).
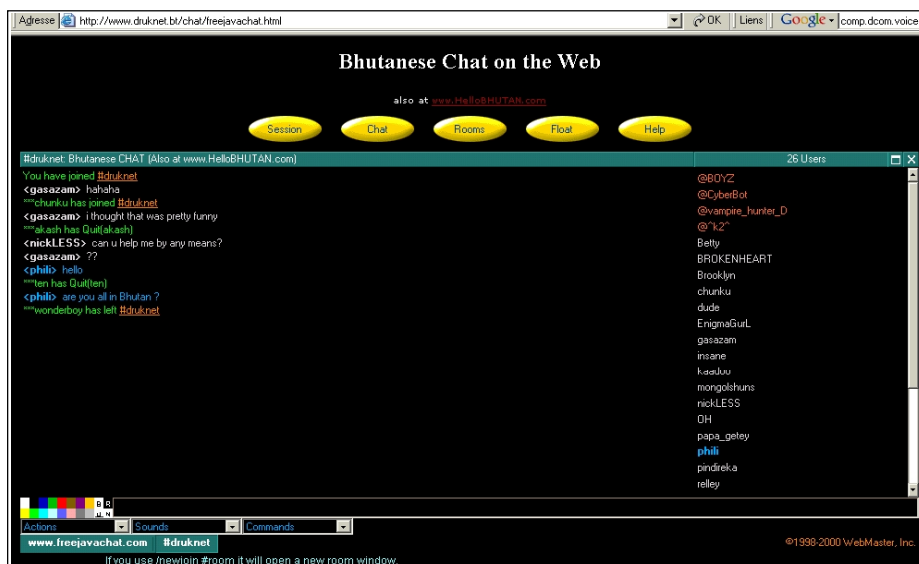


In some websites such as *Verio Inc.*'s (www.verio.com; see picture), a browser-based online chat has been embedded so that users can ask questions and receive answers on the spot. Interestingly, companies such as Nepal-based *Servingminds Pvt. Ltd.*

([www.ServingMinds.com.np](www.ServingMinds.com.np), see section about teleservices in chapter 2) include in their service range the handling of these types of chat-based on-the-spot questions.



**Online collaboration tools** provide advanced functions such as video conference calls, collaboration on shared documents, during-meeting file transfers, use of common electronic whiteboard and text chatting. These tools are traditionally used mostly by professionals and corporations. Well-known online collaboration tools include software such as *Microsoft NetMeeting* (see below) and *Groove Networks ([www.Groove.net](www.Groove.net))*. Interestingly, most people are not aware that *Microsoft NetMeeting* is present in their computers (this program is usually part of the software pre-loaded in a new computer).
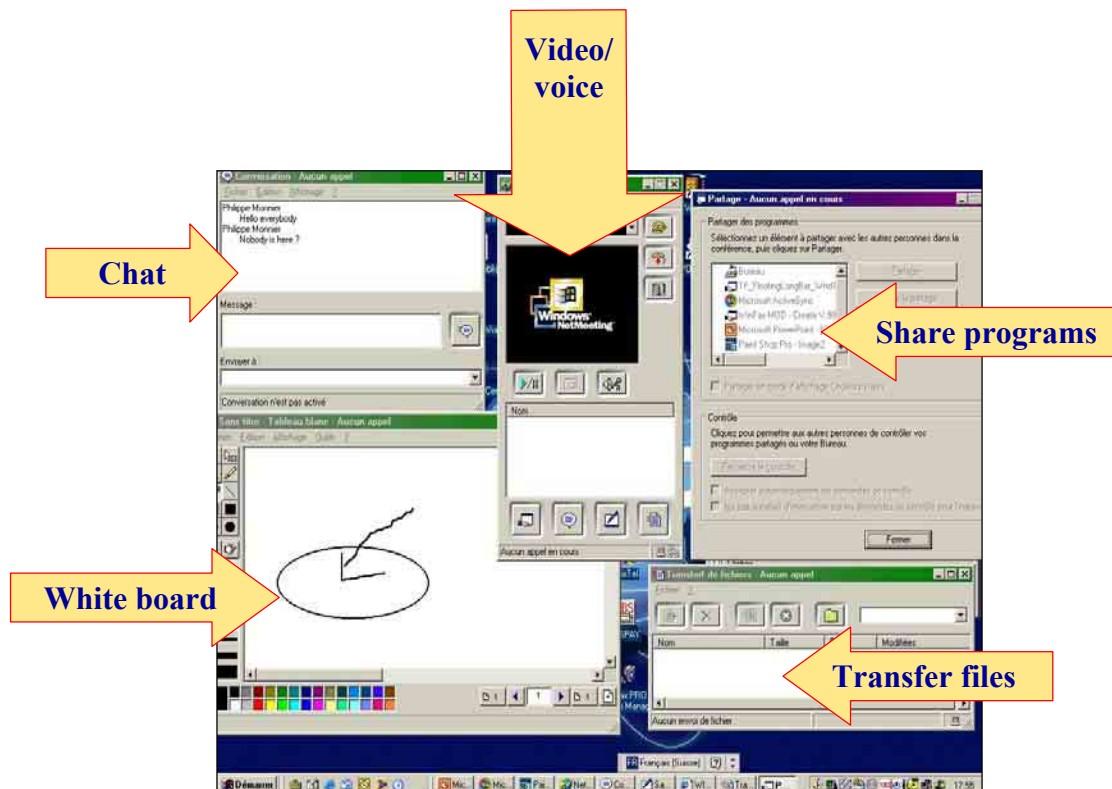
# 13. Understanding Internet search tools

**The Internet has become so enormous that the best search tools cover only a small portion of the Internet.[18] However, the main issue is not the incomplete coverage but the fact that it is difficult to find the most relevant information.**

**Even though the Internet is huge, not everything is available in it**. This is due to at least three reasons:

- <u>Commercial reasons</u>: Information owners may prefer not to have their information loaded in the Internet because they fear that it may be too difficult to sell it online. These same information owners also fear that once their information is on the Internet, it could be downloaded and copied for free (legally or not), making offline sales of the same information much more difficult.
- <u>Practical reasons</u>: Existing information may not be (easily) available in digital format.
- <u>Confidential reasons</u>: Once information has been published on the Internet, it is not possible to fully remove it; indeed, even if some information has been erased from its web server, there are still ways to have access to it (e.g. using a website such as www.archive.org from *Internet Archive*).

**There are thousands of search tools.** They can be differentiated on the basis of the three criteria below:

**(a)** Their **technical coverage** (i.e. which types of files and servers are covered by search tools?)

**(b)** Their **file indexing method** (i.e. are the files to be considered during a search collected by people or by computer programs?)

**(c)** Their **topical coverage** (i.e. do search tools cover very specific topics or a very wide range of topics?)

We will explain in more details these criteria one by one.

**(a) The technical coverage varies very much from one search tool to another.** Potentially, a search tool can search through many file types such as:

- Static (i.e. HTML-based) web pages
- Dynamic/database-generated web pages
- Microsoft Office documents such as Word, Excel, PowerPoint and Access
- Images (gif, jpg, png formats)
- Audio and videos (mp3, mid, wav, formats)
- Acrobat (pdf) documents
- Discussion groups' messages
- Hidden files (i.e. not accessible unless specific actions are taken)
- Files which are no longer online

---

[18] Roughly estimated at about 25 per cent by the author.

Furthermore, **these files can be located in several types of servers,** e.g. web servers FTP servers and surfers' private computers.

What a given search tool covers is constantly evolving because tools tend to improve rapidly even though they still have a lot of difficulties in locating information from dynamic and database-generated webpages. The best way to find out about the status of a particular search tool is to read the details on this matter from the website of each search tool.

**(b) The file indexing method is an essential criterion.** Some search tools are *human-operated*. This means that real people select and classify files (e.g. websites) one by one. As a result, these search tools index a limited number of files but, potentially at least, these files are of high quality. By contrast, in the case of *robot-based* search tools, the indexing is done automatically by computers' programs; consequently, these robot-based search tools have many websites per topic but, on average, websites are of relatively lower quality. Importantly, there is no commonly accepted definition of terms in the field of Internet-based searches. In this manual, when we use the term **search tool** we mean "*any* tool to make Internet-based searches"; when we use the term **search engine**, we mean "robot-based search tools" and when we use the term **directory** we mean "human-operated search tool".

**(c) The third criterion is the topical coverage**. Some search tools concentrate on highly specialized topics (e.g. a given city, a given subject) while others cover everything.

Let's now see a graphical overview of the different types of search tools taking into account the *topical coverage* and the *file indexing method* dimensions. In each category, we have inserted some examples of major search tools; the most influential ones are underlined, i.e. *Google* (market share in 2003: 55 per cent) and *Yahoo* (22 per cent).[19] New terms such as *Meta-Search Tools* and *Expert Sites* will be explained later in this section.

## Figure 1. Types of search tools
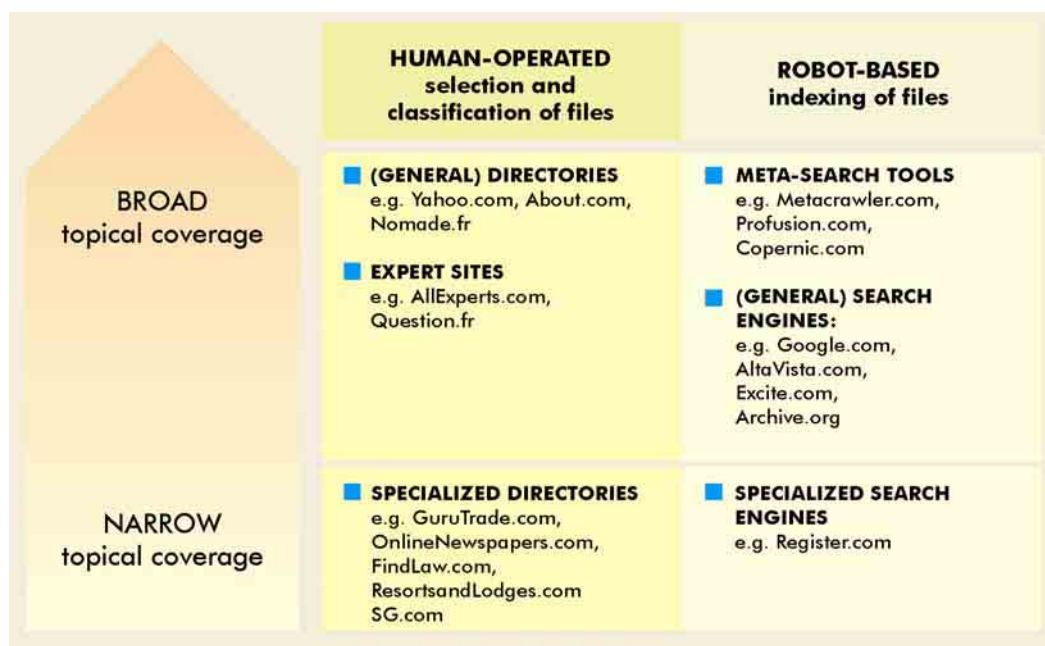


---

Some leading **general (i.e. broad coverage) directories** include:

- *Yahoo! Inc.* ([www.Yahoo.com](www.Yahoo.com)) is clearly the leading directory. Besides its well-known directory, Yahoo also offers an array of services, including a Google-based search engine.
- *About Inc.* ([www.About.com](www.About.com)) is maintained by more than 400 guides living in over 20 countries. Interestingly, guides get 30 per cent of advertising revenues generated by their pages.

Incidentally, to become an *About.com's* guide could be an interesting opportunity for qualified people from developing countries.

**"Expert" sites** provide answers to surfers' questions on the basis of research carried out by teams of experts. Many prominent expert sites used to be free (thanks to the work of many volunteers), but most of them have ceased operation lately. Websites of some leading remaining "expert" sites include:

- [www.ExpertCentral.com](www.ExpertCentral.com) & [www.AllExperts.com](www.AllExperts.com) from *Allexperts.com*: these two pay/free site have joined efforts.
- [www.Exp.com](www.Exp.com) from *EXP Systems, Inc*: this is a pay site
- [www.eQuesto](www.eQuesto): from *eQuesto*: this is a pay site
- [www.Question.fr](www.Question.fr): from *Question.fr Sàrl*: this is a pay site in French only
- [http://answers.google.com/answers/](http://answers.google.com/answers/) from *Google, Inc.*: this is a pay site

Theoretically, expert sites could have either a broad or a narrow coverage. However, most of them (as well as in one of the examples above) have a broad coverage.

As in the case of *About.com's* guides, to become a *Google Answers* researcher could be an interesting opportunity for qualified people from developing countries.

The websites of some leading **specialized (i.e. narrow coverage) directories** include:

- [www.GuruTrade.com](www.GuruTrade.com) (from the *South African Department of Trade and Industry)*, specialized in international trade
- [www.OnlineNewspapers.com](www.OnlineNewspapers.com) (from *onlinenewspapers.org),* specialized in websites of newspapers
- [www.FindLaw.com](www.FindLaw.com) (from *Findlaw*, an independent division of *West Group*), specialized in legal matters
- [www.ResortsandLodges.com](www.ResortsandLodges.com) (from *ResortandLodges.com),* specialized in resorts and lodges
- [http://singaporeeverything.com](http://singaporeeverything.com) (from *SV-Gerald International Pte Ltd*) specialized in Singapore

**Meta-search tools** (also called *meta-search engines*) search using several search tools all at one time. Therefore, their coverage is extremely broad. Leading meta-search tools are quite comparable; websites of these leading meta-search tools include:

- [www.Metacrawler.com](www.Metacrawler.com) from *InfoSpace Inc.*
- [www.Dogpile.com](www.Dogpile.com) from *InfoSpace Inc.*

- www.Profusion.com from *Intelliseek, Inc.*
- www.Metagopher.com from *Metagopher*
- www.Webcrawler.com from *InfoSpace Inc.*
- www.Copernic.com from *Copernic Technologies, Inc.* (a special software, i.e. plug-in needs to be downloaded to use this site)
- www.Kartoo.com from *kartOO* (search results are presented in a map).

The websites of some leading **general (i.e. broad topical coverage) search engines** are as follows:

- www.Google.com (from *Google Inc*) is the clear leader in this field; it will be covered in detail later in this section.
- www.AltaVista.com (from *Overture Services, Inc*.) is the former leader; it has the following functionalities: web, image, MP3/audio, video, Directory, News.
- www.Excite.com (from *The Excite Network*); it has the following functionalities: web, directory, photos, yellow pages, white pages, classifieds.
- www.AlltheWeb.com (from *Overture Services, Inc*.); it has the following functionalities: web, news, pictures, video, audio, FTP files.
- www.Ask.com (from *Ask Jeeves, Inc*.) is a so-called *natural language* search tool (i.e. surfers can ask question in plain English); it has the following functionalities: web, pictures, products, news.
- www.Overture.com (from *Overture Services, Inc*.) ranks search results on the basis of the amount of money paid by website publishers; it covers the web.
- www.Archive.org (from *The Internet Archive*) is a search tool that displays sites as they *used* to be at defined dates.

As a result of keen competition, search engines tend to provide similar functionalities. It should also be noted that the search tool industry is undergoing rapid consolidation. In this regard, *AltaVista* and *AlltheWeb* are now owned by *Overture,* which has been acquired by *Yahoo*.

The websites of some leading **specialized (i.e. narrow topical coverage) search engines** include:

- www.Register.com (from *Register.com*); topical coverage: domain names search/registration
- www.Tile.net (from *SparkNET Corporation*) topical coverage: newsgroups, ftp sites
- www.Topica.com (from *Topica Inc*), formerly www.Liszt.com); main topical coverage: mailing lists (also called newsletters)
- www.Net2one.com (from *Net2One.com*); topical coverage: web clippings
- www.Cyberalert.com (from *CyberAlert, Inc*.) pay site; topical coverage: worldwide
- press clippings from newspapers, magazines, journal, TV

As previously mentioned, specialized search engines do not try to cover all topics and/or all Internet services. Thanks to their specialization, they can – in their area of specialization – perform better than overall leading search tools such as Google. There are obviously many specialized good search engines.

Some search engines are specialized in e-mail addresses (and other personal information). However, these specialized tools tend to be US-oriented and quite incomplete (because people do not like to disclose their e-mail/personal data for fear of being flooded by unwanted promotional e-mails, i.e. spams). Websites of some of these search engines include:

- www.whowhere.lycos.com (from *Lycos, Inc*): e-mail address, telephone number
- www.Bigfoot.com (from *Bigfoot Communications)*: click on *Find People*
- http://People.Yahoo.com (from *Yahoo! Inc*)
- www.WhitePages.com (from *WhitePages.com/W3 Data, Inc*)
- www.Infospace.com (from *InfoSpace Inc*): click on link *World Directories* in the Category *White Pages*

Let's now have a closer look at the clear market leader i.e. **Google** (www.google.com). Google has several major services (**Google Web, Google Images, Google Groups, Google Directory, Google News, Google Language Tools**) and is constantly evolving.

One of the key success factors of Google is its proprietary technology *PageRank*, used to rank search results. In a nutshell, PageRank ranks web pages on the basis of their popularity.[20] The best way to learn how to use Google is to use it as extensively as possible and to refer to the online help features. In the next paragraphs, we will only try to stress some important but less known functions of Google.

**Google Web** is Google's flagship product. Basically, Google Web is a classic search engine with the difference that it uses *PageRank*.



---

[20] In more details, *PageRank* functions as follows, as quoted from *Google.com*. "*PageRank* relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But Google looks at more than the sheer volume of votes, or links that a page receives; it also analyses the page that casts the vote. Votes cast by pages that are themselves 'important' weigh more heavily and help to make other pages 'important' ".

**Google Image** can be used to locate images. For instance, if one's makes a search about His Excellency Samdech Hun Sen, Prime Minister of the Royal Government of Cambodia, using the key phrase *"Hun Sen"*, the results are as follows:



**Google Groups**[21] is an interface to search/view *Usenet* discussion groups. *Usenet* is the largest collection of *discussion groups* (also known as *newsgroups*).

An example of a topic (i.e. trekking in Bhutan) within the newsgroup "rec.climbing" can be found below.



---
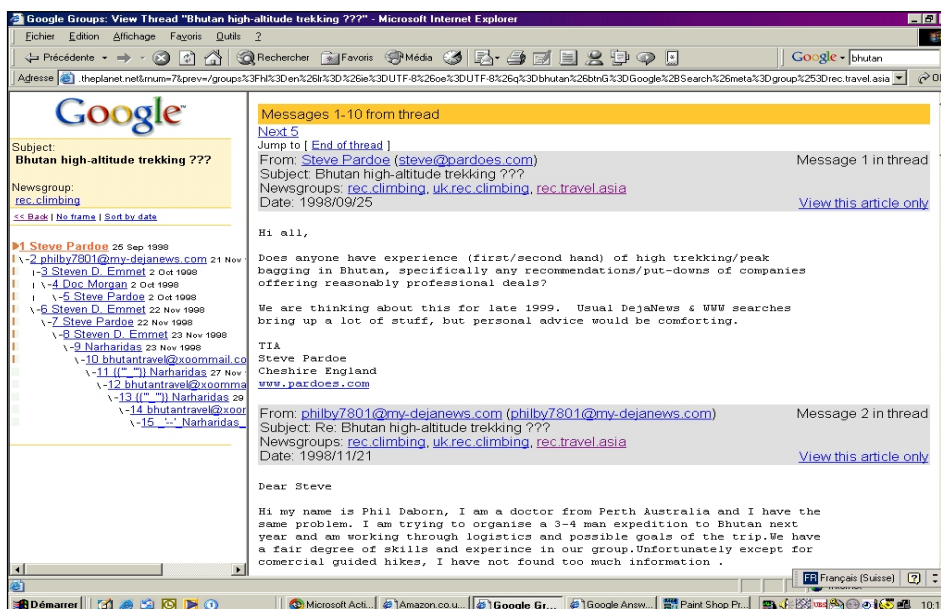
[21] Google Groups is based on www.deja.com, acquired by Google in 2002.

The Google Group interface allows:

- *Message posting* (registration needed)
- *Post nuking* (under certain circumstances). To *post nuke* means to remove a message from a newsgroup
- *Application to create new newsgroups*

**Google Directory:** Even though Google was initially a search engine, it has also a directory (1.5 million web pages are indexed) which is now handled by 20,000 volunteer editors.



**Google News** is based on 4,500 news sources worldwide. It includes articles that appeared within the past 30 days. *Google News* is one of the best ways to search for recent information because it is possible to specify precisely when the news was published (this is not possible using regular search engines, including *Google Web*).
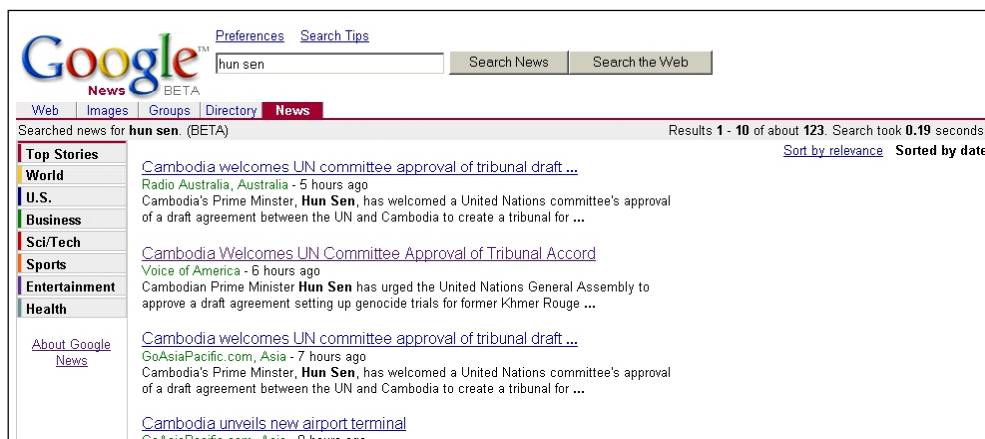


To a large extent, the different types of search tools are complementary. Therefore, **if a complete search is required, it is advisable to use a combination of several tools**. A reasonable approach would include:

- GOOGLE for its overall qualities
- Some SPECIALIZED DIRECTORIES (if any) for their expert human knowledge
- YAHOO & ABOUT.COM because they are leading general directories
- Some META SEARCH TOOLS for their very broad coverage
- SPECIALIZED SEARCH TOOLS, e.g. to find e-mail addresses, domain names

The result of a search is typically a **ranked list** of websites (or webpages). This ranking is very important because surfers tend to consider only the top-ranked results. The main elements considered to rank websites (or webpages) are as follows (in no particular order).

- *URL* (i.e. domain name or website address): Illustration: if a search is made using the keyword "Laos", a website having an URL such as www.laos.com would have a very good chance of being top-ranked.
- *Web page title*
- *Page content* (specially content at the top)
- *" Meta tags"* (e.g. keyword, description, robots: see explanation below)
- *"Alt tags" of images* (i.e. names associated with images; see "technical information" below).
- *Search tool registration details* (i.e. how a website has been exactly registered in a search tool, if applicable)
- *Web page popularit*y (see Google PageRank below)
- *Amount of money* paid to a search tool publisher (e.g. www.Overture.com from *Overture Services, Inc*)

The details of how these ranking criteria work are search-tool-specific, subject to (very) frequent changes and usually not communicated to prevent "search tool spamming" (i.e. the fact of having the same site ranked many times at the top).

## Figure 2. Technical information about meta tags and syntax

**Technical information 1:** Meta tags (not to be confused with meta-search tools) are HTML codes which indicate to search engines what a given webpage is all about. Meta tags are not visible when viewing a website through a browser. In the majority of web pages, webmasters have forgotten to include meta tags.

The syntax of major meta tags is as follows:

```
<HTML>

  <HEAD>

    <TITLE>Your webpage title here </TITLE>

    <META NAME="keywords" CONTENT="keyword1, keyword2, keyword3, ...">
    <META NAME="description" CONTENT="your 200-word description here">
    <META NAME="robots" CONTENT="index, follow">

  <HEAD>
```

Note: users can have a look at the HTML codes of the website www.eyes.ch from eyes.ch to see an intensive use of meta-tags.

Technical information 2: the syntax of ALT tag is as follows (illustrated by an example):
image src="Cambodia_flag.gif" **alt="Flag of Cambodia"**

Website publishers are naturally interested in making sure that many surfers land in their sites through search tools. **To make websites highly visible to search tools**, the above-mentioned ranking criteria must be considered when building a website. In addition, it is possible to use (pay) tools such as *Microsoft's* www.submitit.com or *Netbooster UK Ltd's* www.netbooster.co.uk to have one's website registered in many search tools. The

effectiveness of these popular tools is, however, limited because what matters most is to be extremely well registered in a few prominent search tools; to be so requires manual (as opposed to computer-program-run) actions. One (complex) way to become highly visible to key search tools is to get traffic using a series of so-called satellite sites. The function of these satellite sites is to attract surfers and to redirect them automatically to another site. To maximize the effects of satellite sites, it is important to launch one satellite site for each key search tool; satellite sites must also be updated frequently to ensure that they remain as visible as possible to their respective search tools. As a conclusion, we can state that to be systematically "top-ranked" by key search tools is usually a major and expensive endeavour, especially for sites about popular topics.

It is often important to know **how search tools make money** in order to understand their biases. In other words, it may well happen that some web pages appear at the top of search results, not because their content is outstanding but because they have paid to be top-ranked. Typically, search tools make money using a combination of the following methods:

- Advertisements (text, banner) usually targeted, for instance based on keywords (e.g. *Google, Yahoo*)
- Sale of technology/services (e.g. *Google, AskmeCorp*)
- Fees to  "top-rank" sites (e.g. *Overture*, formerly *AltaVista*)
- Fees to evaluate sites fast (e.g. *Yahoo*: ~US$ 200)
- Any other way to take advantage of well-known brands and surfers' traffic

Finally, a **good reference website to find more information about search tools** is www.SearchEngineWatch.com from *Jupitermedia Corporation*.