

# Internet và tìm kiếm thông tin trên Internet

## **Internet: một thế giới phức tạp**

Internet là chữ viết tắt của “INTer NETwork”, nghĩa là một mạng lưới liên kết với nhau thông qua một giao thức (cách thức giao tiếp) gọi là TCP/IP (*Transmission Control Protocol/Internet Protocol*). Một kiểu hệ thống mạng phổ biến nhất hiện nay là **World Wide Web** (Mạng lưới toàn cầu), kết nối hàng trăm triệu máy tính chủ (*server-computer*) với nhau trên toàn thế giới. Ngoài ra, Internet còn có nhiều công cụ và ứng dụng khác.

## **Các ứng dụng cơ bản trên Internet**

### *Thư điện tử*

(Tiếng Anh: *e-mail*) Ứng dụng này cho phép gửi nhanh chóng một thông điệp cho một hay nhiều người khác nhau có kết nối vào Internet trên khắp thế giới.

Người nhận có thể trích đoạn nội dung thư nhận được, trả lời, đính kèm các tài liệu khác (tập tin, hình ảnh, nhạc,...) và gửi tất cả cho một hoặc nhiều người.

Để gửi được thư điện tử, người gửi và người nhận đều cần có địa chỉ điện tử, thường theo cấu trúc: **tên@têntổchức.tênmiền**. Có nhiều nhà cung cấp dịch vụ thư điện tử miễn phí (và thu lợi nhuận bằng các hình thức khác) như Yahoo, Gmail, Lycos,... Ví dụ: trong địa chỉ *aline2006@yahoo.com* thì *aline2006* = tên, *yahoo* = tên tổ chức, *com* = tên miền.

### *Các danh sách thư điện tử*

(Tiếng Anh: *mailing-list*) Đây là các danh sách dùng cho một nhóm đối tượng nhất định (thường đăng kí miễn phí), để chuyển thông tin theo các chủ đề phù hợp với đối tượng theo kiểu thư điện tử. Thường có các loại danh sách sau:

- **Danh sách thảo luận** (phổ biến nhất): tập hợp các thành viên liên quan trong một chủ đề, cùng thảo luận với nhau về chủ đề đó qua thư điện tử. Danh sách này thường có một người phụ trách việc thêm/bớt thành viên, hoặc lọc thông tin trong các thư trao đổi nếu cần thiết.
- **Các ấn bản định kì** hoặc **bản tin điện tử**: các thành viên trong danh sách sẽ định kì (hàng tuần, hàng tháng,...) nhận được một thư do người quản lí danh sách soạn thảo, và các thành viên không thể trao đổi, thảo luận với nhau.

### *Các diễn đàn thảo luận*

(Tiếng Anh: *newsgroups*) Các diễn đàn này tập hợp các nhóm thảo luận về rất nhiều đề tài khác nhau, thông thường tham gia không cần phải đăng kí.

Các diễn đàn này được chia ra theo nhiều lĩnh vực lớn như: tin học và phần mềm (.**comp**), nghiên cứu và khoa học (.**sci**), kinh doanh (.**biz**), tham khảo tài liệu (.**doc**), giải trí (.**rec**), xã hội (.**soc**),... Các lĩnh vực này lại được chia thành nhiều chủ đề và chủ đề phụ với những nội dung cụ thể hơn, và thường có tên quốc gia đi kèm trước mỗi địa chỉ.

Cấu trúc chung của địa chỉ một diễn đàn thảo luận như sau:

**news://quốcgia.lĩnhvực.chủđề.chủđềphụ**

Ví dụ: khi đăng kí làm thành viên Google Groups, bạn sẽ có nhiều khả năng tham gia nhiều nhóm thảo luận chuyên đề khác nhau.

Tuy cũng tập hợp các bài viết cùng bàn về những chủ đề riêng biệt giống như các danh sách thảo luận, các diễn đàn thảo luận hoạt động theo một cách thức khác: các bài viết không phổ biến cho mỗi người qua thư điện tử mà được tập trung trên một máy chủ (*server*) và cần phải kết nối vào đó mới xem được. Tập hợp các máy tính trao đổi các bài viết dạng này được gọi là **Usenet**. Tuy nhiên, hiện nay các diễn đàn Mạng (*Web forum*, xuất hiện từ những năm 2000) đã dần dần thay thế và chiếm ưu thế.

## *Hội thoại trực tuyến*

(Tiếng Anh: *chat*)

## *Mạng lưới toàn cầu*

(Tiếng Anh: *World Wide Web*) Đây là ứng dụng phổ biến nhất trên Internet, phát triển với tốc độ rất nhanh kể từ khi ra đời vào năm 1989 tại Trung tâm Nghiên cứu Hạt nhân châu Âu (ECNR) có trụ sở tại Geneva, với tác giả là Tim Berners-Lee (tên đầy đủ là Timothy John Berners-Lee, sinh ngày 08 tháng 6 năm 1955 tại Luân Đôn, Anh) .

Một trong những nguyên nhân giúp cho Mạng (*Web*) phát triển rất nhanh là do rất thuận tiện cho người dùng với một hệ thống đa phương tiện (*multimedia*) cho phép hiển thị các trang nhiều màu sắc, có chèn hình ảnh, biểu đồ, các tư liệu âm thanh hay video,... kĩ thuật siêu văn bản (*hypertext*) còn cho phép người dùng dễ dàng di chuyển từ trang này qua trang khác bằng cách nhấp chuột lên các đường siêu liên kết (*hyperlink*). Các trang này có thể nằm trên cùng một điểm mạng (*website*), trên một máy chủ hay ở một quốc gia khác, nên người dùng có thể vượt qua một khoảng cách rất rộng chỉ trong chớp mắt.

Một nét đặc trưng của Mạng là bất cứ người dùng nào cũng có thể vừa là người tiêu dùng vừa là nhà cung cấp thông tin. Viết một trang mạng (*webpage*) khá đơn giản, chỉ cần những phần mềm phổ biến là đủ, và có rất nhiều nhà cung cấp dịch vụ kí gửi điểm mạng miễn phí hay giá rẻ.

Các điểm mạng trên Mạng thường bao gồm nhiều tập tin viết bằng ngôn ngữ đánh dấu siêu văn bản HTML (*Hypertext Markup Language*), đọc được thông qua các

trình duyệt mạng (*web browser*) như Netscape, Internet Explorer, Mozilla Firefox, Opera,... Mỗi tập tin được gọi là một trang mạng, nhưng kích thước không cố định theo kiểu khổ giấy A4 thông thường. Mỗi trang có phần chữ là chủ yếu, và các đường liên kết siêu văn bản (*hypertext link*) hướng đến các trang khác hay các tài liệu có định dạng khác (hình ảnh, âm thanh,...).

## **Phân loại thông tin đăng trên Mạng**

Nhiều người có xu hướng cho rằng “*người ta tìm thấy tất cả trên Mạng*”. Tuy nhiên, cần nói chính xác hơn là “*người ta tìm thấy đủ thứ trên Mạng*”: các ấn bản báo chí; luận văn tốt nghiệp; các trang giới thiệu công sở, cơ quan, doanh nghiệp; các trang viết cảm xúc cá nhân; nhật kí chia sẻ trên Mạng; các diễn đàn trao đổi về mọi vấn đề trên trời dưới đất, từ nghiêm túc đến giải trí, thư giãn,...

*Một nguyên tắc căn bản và thiết yếu nhất trên Internet, đó là phải hậu kiểm thông tin tìm thấy được.* Bất cứ ai cũng có thể viết bất cứ thứ gì để đưa lên Internet, và chính người dùng Internet phải có phận sự kiểm soát thông tin thu nhận được thông qua việc xác định độ tin cậy của nguồn cung cấp, và phải phân biệt giữa thông tin đúng và thông tin sai lệch, vì trong nhiều trường hợp có những thông tin sai lệch được phổ biến một cách không chủ ý.

*Xác định nguồn cung cấp thông tin là cách đơn giản nhất để đánh giá tính phù hợp của thông tin đó.* Để làm điều đó, cần biết phân loại các nguồn thông tin chính trên Internet.

### *Các ngân hàng dữ liệu*

Trong các nguồn thông tin trên Internet, các ngân hàng dữ liệu cung cấp lượng thông tin phong phú vào bậc nhất. Có thể liệt kê:

- Các thông tin tham khảo dạng tóm tắt của các tạp chí chuyên ngành, hội nghị, báo cáo, luận án trong các lĩnh vực khoa học kĩ thuật, y học, khoa học xã hội và nhân văn,...
- Toàn văn của trên 10.000 ấn bản bao gồm nhật báo, tuần báo, tạp chí định kì, ... trong tất cả các lĩnh vực.
- Các thông tin kinh tế dạng toàn văn hay tóm tắt, có chỉ mục (*index*) phân loại chi tiết để giúp tìm kiếm có hiệu quả.
- Các thông tin tham khảo có chỉ mục phân loại hay toàn văn của các bằng sáng chế và thương hiệu đã đăng kí ở các nước phát triển.
- Các danh bạ doanh nghiệp ở hầu hết các quốc gia

Các ngân hàng dữ liệu này tùy loại mà có thể truy cập miễn phí hoặc phải đăng kí thuê bao mới được sử dụng.

## *Các điểm mạng công sở (trường học, chính quyền,...)*

### **Các phòng thí nghiệm và trường đại học**

Đây là các nhà tiên phong của Internet. Các phòng thí nghiệm lớn thường có điểm mạng với rất nhiều nội dung phong phú, bao gồm cả các ngân hàng dữ liệu khoa học, các tài liệu tham khảo, các báo cáo nghiên cứu,... Các trường đại học đa số cũng có điểm mạng riêng, nhưng nội dung thì có sự khác biệt lớn: một số trường chỉ dừng lại ở việc giới thiệu chương trình giảng dạy, cơ sở vật chất, trong khi một số trường khác lại đưa thêm nhiều dịch vụ hỗ trợ về nguồn tài nguyên, tài liệu tham khảo, bài giảng, kết quả nghiên cứu của sinh viên,...

### **Các cơ quan nhà nước**

Ở nhiều nước, các cơ quan nhà nước có điểm mạng riêng và là những nguồn cung cấp thông tin không thể thay thế: số liệu thống kê chính thức, văn bản pháp luật, chủ trương – chính sách của nhà nước, báo cáo, tin tức thời sự,...

### **Các thư viện**

Các thư viện lớn thường có các ngân hàng dữ liệu về thư mục tài liệu lưu trữ trong thư viện. Các ngân hàng dữ liệu này được đưa lên Mạng giúp người đọc tra cứu dễ dàng những thông tin cần thiết trong quá trình tìm kiếm tài liệu tham khảo.

## *Các điểm mạng thương mại*

Khi Internet ra đời, các công ty sản xuất kinh doanh đã nhanh chóng nhận ra lợi ích kinh tế khổng lồ mà hệ thống này mang lại, và cho ra đời các điểm mạng giới thiệu về hoạt động và sản phẩm của mình. Hệ thống điểm mạng thương mại này đã tạo ra một phương thức cạnh tranh mới, và không chỉ vậy mà còn giúp kiểm soát được các hoạt động kinh doanh diễn ra trong từng ngành, lĩnh vực của thị trường. Trên các điểm mạng này có thể cung cấp các thông tin về lịch sử hình thành công ty, danh mục sản phẩm, thông tin tài chính (báo cáo thường niên, báo cáo tài chính...), thông tin về cổ đông, thông cáo báo chí,...

Ngoài ra, có nhiều công ty nhắm đến mục đích trở thành tiêu điểm trong lĩnh vực ngành nghề của mình bằng cách tạo dựng một nguồn tài nguyên thông tin vô cùng phong phú như: tin tức thời sự trong ngành, danh bạ các điểm mạng cùng lĩnh vực, phân tích thị trường,...

## *Các trang cá nhân*

Như đã nói, bất cứ ai cũng có thể đưa thông tin lên Internet. Và trên Mạng có hàng trăm ngàn điểm mạng cá nhân với nội dung và chất lượng rất khác nhau: tự giới thiệu chân dung; đăng hình ảnh kỷ niệm cá nhân, gia đình; chia sẻ sở thích, kinh nghiệm về một thú vui giải trí, một vấn đề cùng quan tâm; sưu tầm các địa chỉ hay; v.v...

Đối với các thông tin do các trang mạng cá nhân cung cấp, cần phải kiểm tra kỹ lưỡng nguồn gốc mới xác định được độ tin cậy của chúng. Có rất nhiều người dùng mạng làm việc hết sức nghiêm túc, tập hợp được nhiều nguồn thông tin có giá trị, và khi tìm kiếm thông tin trên Mạng đây có thể là một điểm xuất phát tốt để rút ngắn thời gian cũng như đào sâu quá trình tìm kiếm. Vấn đề là phải xác định được độ tin cậy của những trang đáng tin cậy, giữa vô vàn những trang không có giá trị cao trong biển thông tin trên Mạng.

## Cách xác định nguồn dữ liệu

Cách xác định nguồn dữ liệu cơ bản nhất là thông qua địa chỉ mạng (*Uniform Resource Locator* – viết tắt là URL – nghĩa là Yếu tố định vị Tài nguyên Đồng dạng). Địa chỉ mạng của một tài liệu sẽ xác định rõ nó thuộc kiểu dịch vụ gì trên Mạng cũng như vị trí của nó trên máy chủ.

Cấu trúc tổng quát của một địa chỉ mạng như sau:

**http://www.tênđiểmmạng.phânloạitênmiền/tênthumục/têntậptin**

Trong đó, các thành phần của địa chỉ mạng bao gồm:

- **http://** - viết tắt của *Hypertext Transfer Protocol* - : xác định kiểu giao thức dùng để truyền thông tin trên Mạng. Thông thường giao thức này là mặc định trong các trình duyệt mạng, do đó có thể không cần nhập thành phần này vào ô địa chỉ trong trình duyệt mà chúng vẫn mở được trang mạng cần đến.
- **www.tênđiểmmạng.phânloạitênmiền**: toàn bộ phần này được gọi là tên miền (*domain name*), do riêng người chủ điểm mạng đăng kí cho phù hợp với nội dung, mục đích, lĩnh vực truyền thông tin của mình (nhưng *vẫn có nhiều trường hợp không kiểm soát được sự lạm dụng*). Tên miền bao gồm các thành phần:
  - **www**: phần này cho phép xác định được ngay đây là một địa chỉ trên Mạng (*Web*). Nhiều trường hợp không bắt buộc phải có phần này trong địa chỉ mạng.
  - **tênđiểmmạng**: do người chủ điểm mạng đăng kí. Tên này có thể thể hiện tên công ti, trường học, công sở, hội nhóm, nhóm sở thích, lĩnh vực, chuyên môn,... Tên này thông thường không mang dấu, không có khoảng trắng và một số kí tự đặc biệt. Hiện có một số nhóm vận động sử dụng tên điểm mạng có dấu, nhưng vẫn chưa được hưởng ứng rộng rãi.
  - **phânloạitênmiền**: có thể có nhiều cấp phân loại tên miền, mỗi cấp cách nhau bằng một dấu chấm, cấp càng cao càng đứng phía sau. Cấp phân loại tên miền cao nhất là tên miền cấp 1 (*Top Level Domain*). Tên miền cấp 1 có hai loại là tên miền cấp 1 chung (*generic Top Level Domain*, viết tắt **gTLD**) để chỉ lĩnh vực hoạt động và tên miền cấp 1 mã quốc gia

(*country codes Top Level Domain*, viết tắt **ccTLD**) để chỉ giới hạn địa lí. Tên miền loại **gTLD** có một số loại phổ biến sau:

<b>gTLD</b>	<b>Ý nghĩa</b>	<b>Đối tượng đăng kí</b>
.aero	Hàng không	Các đối tượng thuộc ngành hàng không
.biz	Kinh doanh	Tất cả (trừ một số trường hợp đặc biệt)
.com	Thương mại	Tất cả
.edu	Giáo dục	Các trường đại học, cao đẳng
.gov	Chính phủ	Các cơ quan nhà nước Hoa Kỳ
.info	Thông tin	Tất cả
.int	Quốc tế	Các tổ chức thành lập thông qua các hiệp ước quốc tế
.name	Tên họ	Cá nhân, có thể dành cho pháp nhân bảo vệ thương hiệu
.net	Mạng lưới	Tất cả
.mil	Quân sự	Các cơ quan quân sự Hoa Kỳ
.museum	Bảo tàng	Các viện bảo tàng đáp ứng yêu cầu của Hội đồng các Viện bảo tàng Quốc tế (ICOM)
.org	Tổ chức, liên minh	Tất cả
.pro	Chuyên gia nghề nghiệp	Các luật sư, bác sĩ hay chuyên gia nghề nghiệp khác

Các **gTLD** do InterNIC quản lí và điều phối, vẫn có thể được tạo mới khi có yêu cầu phù hợp. Một số **gTLD** như **.com**, **.biz**, **.info**, **.net**, **.org** được tạo ra với mục đích ban đầu là dành riêng cho các lĩnh vực hoạt động tương ứng, nhưng vẫn có hiện tượng lạm dụng ít hay nhiều, nội dung thông tin của điểm mạng không phù hợp với lĩnh vực đã đăng kí, có khi nhằm mục đích che mắt, đánh lừa người dùng mạng.

Tên miền loại **ccTLD** bao gồm 2 chữ cái, viết tắt tên quốc gia: **.us** (Hoa Kỳ), **.uk** (Vương quốc Anh), **.fr** (Pháp), **.ca** (Canada), **.vn** (Việt Nam), **.cn** (Trung Quốc), **.jp** (Nhật Bản), **.kr** (Hàn Quốc),... Tên miền của quốc gia nào do quốc gia đó quản lí, và cấp cho các điểm mạng muốn thể hiện “*thuộc quốc gia đó*”. Các tên miền cấp 2 dưới cấp **ccTLD** do mỗi quốc gia định nghĩa và quản lí, có thể là tên của chủ điểm mạng (công ti, công sở, pháp nhân) hoặc lĩnh vực hoạt động (**.com/.co** cho thương mại, **.edu/.ac** cho giáo dục, **.net** cho mạng lưới, **.org** cho tổ chức,...) Mỗi chủ sở hữu của một tên miền có quyền tạo ra các tên miền cấp dưới, trong địa chỉ mạng sẽ *đứng trước cấp thấp nhất trong phần đăng kí tên miền chính thức* của mình. Ví dụ: Trường Đại học Khoa học Tự nhiên TP.HCM có tên miền cấp 3 là **hcmuns** (hcmuns.edu.vn), có quyền cấp tên miền cấp 4 cho Khoa Công nghệ Thông tin là **fit** (fit.hcmuns.edu.vn).

Mỗi điểm mạng đều có một trang chủ để tiếp đón người dùng mạng, ở đó thường có các thông tin giới thiệu tổng quát và các đường siêu liên kết đến các phần nội dung, tài liệu khác nhau bên trong điểm mạng. Trang này thường được đặt tên là **index** hoặc **default** hay **home**. Khi nhập tên miền của một điểm mạng vào ô địa chỉ của trình duyệt mạng, trình duyệt sẽ tự động tìm đến trang chủ này.

- **Thư mục:** mỗi điểm mạng thường được chia thành nhiều thư mục hoặc cấp thư mục khác nhau dưới cấp thư mục gốc nhằm giúp việc sắp xếp, quản lí tài liệu và các trang mạng một cách khoa học, trật tự và gọn gàng. Các thư mục này cũng là thành phần của địa chỉ mạng, và thông thường không dùng kí tự có dấu, khoảng trắng hay kí tự đặc biệt nhằm tránh các sự cố khi truy cập thông qua các trình duyệt mạng.
- **Tập tin:** mỗi trang mạng thể hiện dưới dạng một tập tin. Ngôn ngữ phổ biến để tạo một tập tin trang mạng là HTML. Thông qua trình duyệt mạng, có thể mở một trang mạng bằng cách nhấp chuột vào đường siêu liên kết dẫn chính xác đến tập tin của trang đó, hoặc gõ trực tiếp địa chỉ đầy đủ của trang đó (bao gồm cả các thư mục, thư mục con và phần mở rộng của tập tin).

Với những hiểu biết cơ bản về tên miền, người dùng mạng có thể xác định được nguồn gốc của nhiều thông tin, tài liệu tìm được. Thông thường, ngoài các yếu tố về chuyên môn, *tên miền càng rõ ràng về nguồn gốc thì thông tin được cung cấp có độ xác thực càng cao, tên miền của các tổ chức, cơ quan có tính pháp lí càng cao thì độ tin cậy của thông tin bên trong đó càng cao.*