Lời cảm ơn

Trong thời gian thực hiện đồ án "Nghiên cứu một số vấn đề về khái thác và tìm kiếm dữ liệu trên cổng thông tin điện tử " Em đã nhận được sự hướng dẫn ,chỉ bảo và giúp đỡ tận tình của các thầy ,cô khoa công nghệ thông tin trường Đại Học Dân Lập Hải Phòng .Vậy cho phép em được bày tỏ lòng biết ơn sâu sắc tới sự giúp đỡ đó.Đặc biệt em xin chân thành cảm ơn Thầy giáo -Thạc sĩ: Võ Văn Tùng -Người đã trực tiếp hướng dẫn và tạo mọi điều kiện thuận lợi giúp đỡ em hoàn thành đồ án này .Qua đây em cũng xin cảm ơn gia đình ,bạn bè đã giúp đỡ và động viên em hoàn thành đồ án này

Vì thời gian có hạn, trình độ bản thân còn nhiều hạn chế. Cho nên trong đề tài không tránh khỏi những thiếu sót, em rất mong được sự góp ý quý báu của tất cả các thầy cô giáo cũng như các bạn để đề tài của em được hoàn thiện hơn.

Em xin chân thành cảm ơn !

Hải Phòng, tháng 7 năm 2009

Sinh viên

PHẦN MỞ ĐẦU

Trong những năm gần đây, các ứng dụng trên Internet phát triển nhanh, ảnh hưởng của nó là đã làm thay đổi nhiều đến đời sống kinh tế, văn hoá, xã hội của tất cả các nước trên thế giới. Trong sự phát triển mạnh mẽ của Internet, thì các Website giữ một vai trò đặc biệt quan trọng. Tuy nhiên, với thời gian hình thức này đã bộc lộ một số nhược điểm cần phải khắc phục. Cùng với sự trợ giúp của công nghệ Soft Agent - một chương trình thay mặt người dùng thực hiện công việc tìm kiếm và xử lý thông tin trên Internet - khái niệm Website truyền thống được chuyển thành "Website thông minh" với sự trợ giúp của dịch vụ Search Engine, một công cụ cho phép tìm kiếm và lọc thông tin trên cơ sở các từ khoá được xác lập bởi người dùng và dịch vụ phân loại thông tin – Category. Từ đó, thuật ngữ "Website thông minh" hay "Cổng thông tin điện tử" - Portal được hình thành.

Hiện nay, một số quốc gia, một số tổ chức trên thế giới đã quan tâm chú ý đến sự phát triển công nghệ Portal, công nghệ này đã và đang trở thành xu thế chung trong quá trình phát triển trên Internet. Ở nước ta, một số địa phương cũng rất quan tâm phát triển công nghệ Portal như thành phố Hà Nội, thành phố Hồ Chí Minh, tỉnh Hà Tây và một số địa phương khác… Các địa phương này đã xây dựng được cổng thông tin điện tử cho riêng mình, nó đã trở thành một công cụ phục vụ đắc lực trong việc quản lý, điều hành các hoạt động kinh tế, xã hội.

Cũng như một số ngành kinh tế - xã hội khác, ngành Giáo dục và Đào tạo với đặc điểm quản lý một địa bàn trên diện tích rất rộng lớn, việc tổng hợp phân tích các số liệu có liên quan đến hoạt động của ngành ở các địa phương tại các thời điểm khác nhau là rất cần thiết để Bộ Giáo dục và Đào tạo có thể đưa ra các biện pháp điều chỉnh đúng đắn và kịp thời. Chính vì vậy, việc xây dựng nghiên cứu thiết kế và tổ chức dữ liệu trên cổng thông tin điện tử ngành Giáo dục và Đào tạo để phục vụ cho công tác quản lý, chỉ đạo chuyên môn là việc làm cần thiết, góp phần vào việc đổi mới, nâng cao chất lượng Giáo dục và Đào tạo.

Xuất phát từ nhu cầu trên, em hướng nghiên cứu của mình vào các vấn đề liên quan đến lĩnh vực tổ chức dữ liệu và các giải pháp kỹ thuật hỗ trợ khai thác và tìm kiếm dữ liệu trên cổng thông tin điện tử. Về kết cấu của luận văn, ngoài phần mở đầu, kết luận và tài liệu tham khảo, luận văn được trình bày trong 3 chương:

Chương 1: Tổng quan về cổng thông tin điện tử Portal

Nội dung chương trình bày tổng quan về Portal.

Chương 2:Nghiên cứu một số vấn đề về tổ chức dữ liệu, cơ chế chuyển đổi dữ liệu trong cổng thông tin phục vụ cho việc tìm kiếm và khai thác dữ liệu.

Tìm hiểu tổ chức CSDL trong hệ thống thông tin phân tán; nghiên cứu một số phương pháp tìm kiếm và khai thác dữ liệu trên cổng thông tin điện tử iết lập cơ chế chuyển đổi thông tin tự động giữa các sever; Một số giải thuật tìm kiếm thông tin trên hệ thống thông tin phân tán.

Chương 3: Áp dụng nghiên cứu chương trình giải quyết bài toán khai thác và tìm kiếm thông tin trên cổng thông tin của ngành Giáo dục và Đào tạo

Trong chương này, trên cơ sở nghiên cứu và phân tích các yêu cầu thực tế từ các đơn vị, đưa ra các chuẩn hoá dữ liệu, thiết kế xây dựng cổng thông tin giáo dục và hướng giải quyết bài toán khai thác, tìm kiếm thông tin trong Cổng thông tin giáo dục.

Chương 1
TỔNG QUAN VỀ CỔNG THÔNG TIN ĐIỆN TỬ

1.1.Khái niêm về portal

1.1.1.Định nghĩa portal

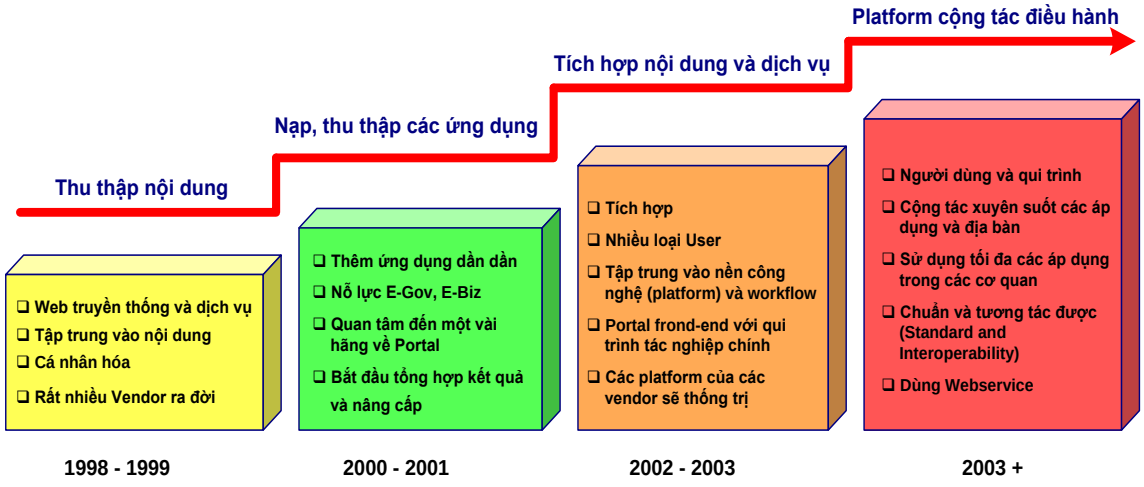
a.Lịch sử cổng thông tin điện tử

Website đã và đang đóng góp rất lớn vào việc phổ cập thông tin, như :giới thiệu tin tức, các cơ sở dữ liệu, và một số chương trình ứng dụng trên mạng, đã làm thay đổi cả thế giới từ khi xuất hiện vào đầu những năm 90 của thế kỷ trước. Ngày nay mọi giao dịch thông qua web đã trở nên phổ biến.

Công nghệ Portal (Cổng điện tử) phát triển sau thời kỳ này khoảng 7-8 năm như là một tất yếu xuất phát từ nhu cầu thực tế. Portal là một bước tiến hóa của web truyền thống. Nó ra đời để giải quyết những vấn đề mà website truyền thống gặp phải.

- Portal (cổng giao tiếp điện tử) là một bước tiến hóa của website truyền thống.
- Là "siêu website", gọi đầy đủ là Portal Website, gọi tắt là portal, đối với người dùng vẫn chỉ là sử dụng trang web thông qua trình duyệt (tức là web browser), nhưng đằng sau nó là sự thay đổi thuật ngữ và quan niệm mới về triết lý phục vụ thay cho cách hiểu "tuyên truyền" thông qua website như trước đây.
- Là điểm đích quy tụ hầu hết các thông tin và dịch vụ cho người sử dụng cần, là điểm đích đến thực sự. Thông tin và dịch vụ được phân loại nhằm thuận tiện cho tìm kiếm và hạn chế vùi lấp các thông tin.
- Bảo toàn đầu tư lâu dài. Có nền tảng công nghệ đảm bảo, do công nghệ Internet đã phát triển rất cao so với thời kỳ xuất hiện Word Wide Web vào đầu những năm 90 của thế kỷ trước. Những công nghệ tạo nên thời đại portal đều hỗ trợ tính mở và kế thừa rất mạnh, sao cho việc mở rộng quy mô phục vụ bằng các phần mềm ứng dụng mới được "lắp rắp" vào Portal đang có mà không phải hủy bỏ hoặc sửa chữa lớn như những website trước đây.
- Môi trường chủ động dùng cho việc tích hợp ứng dụng.

- Xu hướng "tiến hóa" chung của website theo hướng tiến đến portal được trình bày trong hình vẽ:



Phát triển của Portal

b.Cổng thông tin điện tử -Portal là gì?

Portal hay Cổng thông tin điện tử được hiểu như là một trang web xuất phát mà từ đó người sử dụng có thể dễ dàng truy xuất các trang web và các dịch vụ thông tin khác trên mạng máy tính. Ban đầu khái niệm này được dùng để mô tả các trang web khổng lồ như là Yahoo, Lycos, Altavista, AOL… bởi mỗi ngày có hàng trăm triệu người sử dụng chúng như là điểm bắt đầu cho hành trình "lướt web" của họ. Lợi ích lớn nhất mà portal đem lại là tính tiện lợi, dễ sử dụng. Thay vì phải nhớ vô số các địa chỉ khác nhau cho các mục đích sử dụng khác nhau, thì với một web portal như Yahoo, người dùng chỉ cần nhớ yahoo.com, ở trong đó nhà cung cấp dịch vụ đã tích hợp mọi thứ mà khách hàng cần…

- Bạn bắt đầu hành trình "lướt web" của mình như thế nào?
- Yahoo
- Tại sao lại là Yahoo?
- Đó là một trang Web cho phép ta dễ dàng truy nhập tới mọi thứ ta cần: tìm kiếm thông tin, đọc tin tức, tán gẫu với bạn bè, gửi thiệp, gửi thư điện tử, xem giá chứng khoán,

thậm chí mua sắm một thứ gì đó.

- Đúng thế, có rất nhiều trang web như vậy trên mạng, người ta thường gọi chúng là các portal.

Với các đặc tính như 'chỉ một kết nối' hay 'tất cả trong một' các web portal đã trở thành một đầu mối thông tin cho mọi vấn đề, một thứ la bàn định hướng cho người dùng trong hành trình khám phá kho báu internet rộng lớn.

Ngày nay khái niệm portal không chỉ áp dụng cho các 'gã khổng lồ truyền thông' kể trên, nguyên lý một đầu mối cho tất cả đã được áp dụng vào việc nâng cấp, cải tạo các website kiểu cũ, góp phần hình thành nên một không gian portal (portal space) trên mạng internet.

Các nhà cung cấp dịch vụ internet (ISP) xây dựng nên các portal để hỗ trợ khách hàng của mình trong việc sử dụng internet. Các dịch vụ mà họ thường tích hợp vào trong portal của mình là công cụ tìm kiếm, danh mục các trang web được sắp xếp theo một tiêu chí nào đó, trang tin tức điện tử, dịch vụ nhắn tin, phòng chat, hòm thư điện tử hay trang web cá nhân miễn phí … Các portal này cố gắng để tạo ra một thế giới internet thu nhỏ cho các khách hàng, vì thế chúng thường được khuyến cáo như là điểm bắt đầu lý tưởng cho những người mới tìm hiểu về internet.

Khác với mục đích xây dựng portal bao trùm mọi lĩnh vực mà các công ty truyền thông theo đuổi, những cộng đồng chuyên môn trên mạng Internet chỉ muốn xây dựng portal phục vụ cho duy nhất một lĩnh vực mà mình quan tâm. Vẫn với nguyên lý 'một đầu mối cho tất cả', các portal này thường đi sâu vào nghiên cứu nhiều khía cạnh khác nhau của một vấn đề. Người ta gọi chúng là các portal chuyên môn hay vortal (vertical portal).

Sức hấp dẫn của các portal không chỉ bởi sự tập trung thông tin về một đầu mối, chúng còn có một tính năng quan trọng khác đó là khả năng tương tác thông tin nhiều chiều. Nói một cách khác đi, người dùng không chỉ khai thác thông tin từ portal mà họ còn có thể đưa ra những yêu cầu để được phục vụ. Các portal được xây dựng cho chính phủ, cho chính quyền tỉnh, thành phố là một ví dụ. Ngoài vai trò như một 'tổng hành dinh trực tuyến' nơi đóng quân của đầy đủ các sở ban ngành, các portal này còn cho phép

người dân làm những việc như đăng ký kinh doanh qua mạng, đăng ký kết hôn qua mạng… thậm chí bỏ phiếu bầu cử qua mạng. Mọi đối tượng sử dụng đều có thể tìm kiếm và khai thác kho thông tin đa dạng này một cách dễ dàng qua một giao diện thống nhất mà không cần biết thông tin này ở đâu, do ai quản lý. Chẳng hạn, người dân có thể tìm thấy và sử dụng ngay dịch vụ hành chính mà họ cần, chứ không phải quan tâm đến cấp chính quyền nào, những cơ quan nào liên quan đến các thủ tục đó.

Song song với sự phát triển của các portal như Yahoo, AOL… Các tập đoàn công nghệ thông tin lớn cũng sử dụng cách tương tự để cải tiến hệ thống thông tin của mình. Họ đã tạo ra những mô hình kiểu mẫu cho việc xây dựng các portal doanh nghiệp (EIP-Enterprise Information Portal). Các portal như thế này trước hết là để phục vụ cho các công việc của doanh nghiệp, mà cụ thể là hỗ trợ các tiến trình truyền thông và tương tác giữa các cá nhân, bộ phận trong doanh nghiệp (B2E – Business to Employee). Một số mô hình EIP của mạng thông tin nội bộ (Business Intranet Portal) cho phép các nhân viên dễ dàng khai thác các nguồn tài nguyên thông tin trong doanh nghiệp đồng thời cho phép truy xuất ra các portal công cộng, các portal chuyên ngành hẹp khác. Portal cộng tác, tạo một môi trường làm việc ảo cho phép các nhân viên có thể làm việc với nhau từ bất cứ đâu. Portal chuyên gia, kết nối các nhân viên dựa trên yếu tố năng lực của từng người… Các ứng dụng đa dạng của portal trong môi trường nội bộ doanh nghiệp là một công cụ không thể thiếu đối với các doanh nghiệp trong thời đại bùng nổ thông tin, đặc biệt là đối với những doanh nghiệp có nhiều bộ phận, chi nhánh phân bố trong một không gian địa lý rộng. Cũng vẫn trong môi trường ứng dụng là các doanh nghiệp, công nghệ portal còn cung cấp một công cụ giao tiếp hữu hiệu với thế giới bên ngoài. Khái niệm cổng thông tin doanh nghiệp mở rộng (Extended enterprise portal - extranet) nhằm nói tới một trang web cho phép doanh nghiệp thực hiện giao dịch với các khách hàng của mình (B2C) hay với các nhà cung cấp, các đối tác (B2B).

Các doanh nghiệp nhỏ khó có thể tự xây dựng cho mình một portal đầy đủ tiêu chuẩn, tuy nhiên nếu muốn họ vẫn có thể tiến hành các giao dịch qua mạng thông qua các chợ điện tử (e-Marketplace portal). Chợ điện tử là một portal về xúc tiến thương mại, các

doanh nghiệp tham gia chợ điện tử như thể tham gia một kỳ triển lãm. Ở đó, các doanh nghiệp có thể tiếp cận nguồn thông tin về thị trường, gặp gỡ các khách hàng tiềm năng, các đối tác…

Vai trò của portal là không thể phủ nhận đối với các hoạt động trên mạng internet, . Tuy nhiên cũng cần phải khẳng định rằng việc xây dựng một portal thực thụ là việc không đơn giản.Ở đây em chỉ đi vào nghiên cứu thế nào là một portal và tìm hiểu một số vấn đề về khai thác và tìm kiếm thông tin thông qua cổng thông tin điện tử .Qua đó áp dụng vào việc khai thác và tìm kiếm thông tin trên cổng thông tin của bộ giáo dục và đào tạo

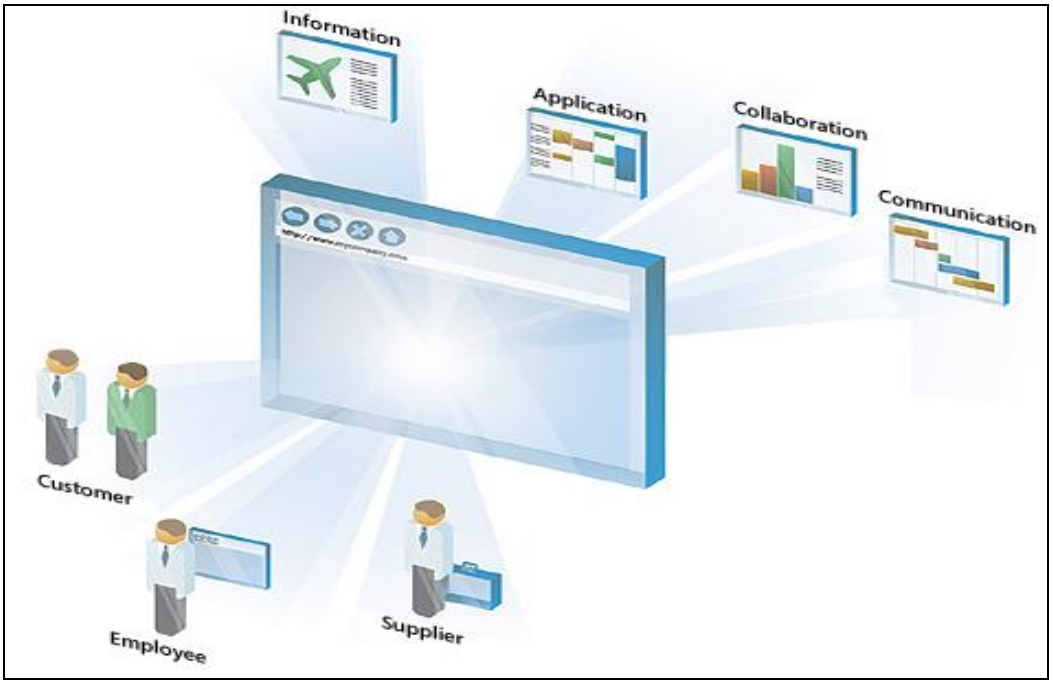
,

c.Định nghĩa:

Cổng thông tin điện tử - Portal: là một khái niệm thường được nhắc đến nhiều trong những năm gần đây của thị trường tin học. Bởi vì phạm vi áp dụng của Portal là rất rộng, bao gồm các hệ thống bên trong (internal), bên ngoài (external), đằng sau bức tường lửa và nằm rải rác khắp nơi trên internet, do vậy ta khó có được định nghĩa hoàn chỉnh và chính xác về Portal. Một cách chung nhất,ta có thể tạm định nghĩa portal như sau:

- Portal là giao diện dựa trên nền web được tích hợp và cá nhân hóa tới các thông tin, ứng dụng và các dịch vụ hợp tác .
- Portal như là một cổng tới các trang web, cho phép một khối lượng lớn các thông tin sẵn có trên Internet và các ứng dụng được tích hợp, được tuỳ biến, được cá nhân hóa theo mục đích của người sử dụng .
- Portal là điểm đích truy cập trên Internet mà qua đó người dùng có thể khai thác mọi dịch vụ cần thiết và “không cần thiết phải đi đâu nữa”.
- Protal là một giao diện web đơn,nó cung cấp truy cập cá nhân tới thông tin ,các ứng dụng ,xử lí thương mại và nhiều hơn nữa . Với công nghệ Portal, các tổ chức có thể giảm cường độ, nhưng lại tăng giá trị lao động và đặc biệt còn làm tăng giá trị các sản phẩm. Các tổ chức có thể tích hợp thông tin trong phạm vi môi trường làm việc, các ứng dụng dịch vụ hoặc sử dụng giao diện đơn lẻ .
- Portal là một giao diện dựa trên nền Web, tích hợp các thông tin và dịch vụ có thể có. Nó cho phép khai báo, cá biệt hóa thông tin và dịch vụ, cho phép quản

trị nội dung và hỗ trợ một chuẩn về một nội dung và giao diện hiển thị. Nó cung cấp cho người dùng một điểm truy cập cá nhân, bảo mật tương tác với nhiều loại thông tin, dữ liệu và các dịch rộng rãi đa dạng ở mọi lúc mọi nơi nhờ sử dụng một thiết bị truy cập Web



Hình ảnh về một portal

1.1.2.So sánh portal với một website thông thường

a.Bảng so sánh portal với website thông thường

Portal	**Website thông thường**
+ Portal hỗ trợ khả năng đăng nhập một lần tới tất cả các tài nguyên được liên kết với Portal. Nghĩa là, người dùng chỉ cần một lần đăng nhập là có thể vào và sử dụng tất cả các ứng dụng đã được tích hợp trong Portal đó mà người dùng này có quyền.	Một website thông thường không có được khả năng đăng nhập một lần.
+ Portal hỗ trợ khả năng cá nhân hóa	Thường không hỗ trợ, nếu có chỉ ở mức

theo người sử dụng. Đây là một trong những khả năng quan trọng của Portal, giúp nó phân biệt với một website thông thường. Portal cá nhân hóa nội dung hiển thị, thông thường đây là sự lựa chọn một cách tự động dựa trên các quy tắc tác nghiệp, chẳng hạn như vai trò của người sử dụng trong một tổ chức. Ví dụ khi một người mua hàng đăng nhập vào hệ thống, Portal sẽ hiện ra một danh sách các sản phẩm mới. Hoặc nếu cần quan tâm đến các lĩnh vực khảo cổ thì Portal có thể cung cấp các thông tin bảng danh sách các đồ cổ.	độ rất nhỏ, không phải là đặc điểm nổi bật.
+ Khả năng tùy biến. Đây là một khả năng tiêu biểu của một Portal. Ví dụ một giao diện Portal có mục thông tin thời tiết, chúng ta có thể bỏ phần thông tin này đi nếu chúng ta không quan tâm đến nó. Hoặc chúng ta có thể thay đổi cách hiển thị của Portal. Ví dụ như thay vì hiển thị bằng font chữ màu xác định chúng ta có thể thay nó bằng chữ màu đỏ, hay có thể tự thay đổi giao diện của Portal nếu mặc định chức năng A được đặt sau chức năng B, nếu	Một vài Website có nhưng chỉ dừng lại ở mức độ dựng sẵn, người dùng chỉ có thể lựa chọn một vài giao diện đã có, mà không tự mình thay đổi từng mục một cách tùy ý.

không thích chúng ta có thể thay đổi lại thứ tự hiển thị này. Đặc tính này tương tự như màn hình desktop của chúng ta.	
+ Liên kết truy cập tới hàng trăm kiểu dữ liệu, kho dữ liệu, kể cả dữ liệu tổng hợp hay đã phân loại. Portal nó có khả năng liên kết tới tài nguyên dữ liệu rộng lớn, gồm nhiều kiểu dữ liệu từ dữ liệu thông thường đến siêu dữ liệu.	Chỉ sử dụng các liên kết để tới các site khác nhưng nội dung chủ yếu vẫn chỉ tập trung trong trang đó.
+ Portal hỗ trợ rất tốt khả năng liên kết và hợp tác người dùng. Portal không chỉ liên kết chúng ta với những gì chúng ta cần mà còn liên kết với những người mà chúng ta cần. Khả năng liên kết này được thực hiện bởi các dịch vụ hợp tác.	- Không hỗ trợ

Trên đây là những so sanh để thấy được sự khác nhau của một Portal với những trang web thông thường .Tuy nhiên Hiện tại trên thị trường có khá nhiều giải pháp hoặc sản phẩm portal, mỗi sản phẩm có một sắc thái riêng, sử dụng công nghệ riêng, phục vụ cho đối tượng riêng, ... và vô hình chung sự "đa dạng" này dẫn tới tình trạng khó chọn lựa một giải pháp phù hợp với nhu cầu cụ thể. Vì vậy, để phân biệt giữa giải pháp portal với một ứng dụng web hay một phần mềm quản trị nội dung, bạn phải lựa chọn giải pháp phù hợp của nhiều nhà cung cấp, đảm bảo việc ứng dụng CNTT với portal là đúng hướng, mà không giới hạn portal phải theo một công nghệ nào.

b.Các bước so sánh portal với website thông thường

Khả năng cá nhân hoá (Personalization)

Để đánh giá tính năng này, bạn cần yêu cầu nhà cung cấp trình diễn hoặc giới thiệu cách thức hệ thống cung cấp thông tin cho nhiều người dùng khác nhau hoặc nhiều cấp độ

người dùng khác nhau. Tại đây có thể có nhiều kết quả khác nhau. Nếu với 2 người dùng khác nhau hoặc với 2 cấp độ sử dụng (quyền) khác nhau và thông tin hiển thị vẫn giống nhau, thì bạn có thể kết luận ngay rằng hệ thống này không có phép cá nhân hoá thông tin, và có thể đi đến kết luận cuối cùng rằng đó không phải là hệ thống portal. Nếu với 2 cấp độ khác nhau, thông tin được sử dụng có sự khác nhau thì có thể đi đến kết luận hệ thống này cho phép cá nhân hoá thông tin theo thẩm quyền sử dụng.

Khả năng tích hợp nhiều loại thông tin (Content aggregation)

Đây là một đặc tính quan trọng bậc nhất của hệ thống portal, đặc tính này thể hiện portal có thể mở rộng được hay không. Đặc tính này thể hiện qua thuật ngữ "ghép là chạy", có nghĩa là khi cần mở rộng thêm thành phần (module) dịch vụ mới, thì chỉ cần điều chỉnh và tích hợp lại thông tin của module dịch vụ đó một cách đơn giản, nhanh chóng và tức thì đối với hệ thống mà không phải biên dịch lại hoặc viết lại mã chương trình.

Để kiểm định tính năng này, bạn hãy yêu cầu nhà cung cấp trình diễn hoặc giới thiệu cách thức hệ thống tích hợp thông tin từ nhiều module dịch vụ khác nhau của hệ thống, ví dụ như hiển thị một nội dung bài viết trong một màn hình, bên cạnh đó là danh sách các chủ đề thảo luận trong forum. Tại đây có thể có nhiều kết quả khác nhau.

+Nếu nhà cung cấp khi bổ sung ứng dụng/dịch vụ vào portal mà phải “bẻ” mã (code) của website ra để viết thêm module về màn hình, các liên kết trang, các truy cập cơ sở dữ liệu mới, một hệ thống phân quyền sử dụng mới, v.v... thì hệ thống đó không gọi là có tính mở được, vậy kết luận là hệ thống không có khả năng tích hợp ứng dụng theo kiểu “ghép là chạy”, và có thể kết luận ngay hệ thống đó không phải là giải pháp portal.

+ Nếu hệ thống cho phép "ghép" các ứng dụng lại với nhau, bạn hãy yêu cầu nhà cung cấp thay đổi nguồn hoặc kênh thông tin của các ứng dụng đã tích hợp, nếu không thế thì kết luận "đó là hệ thống giả portal" chứ không phải là giải pháp portal.

+ Nếu có thể tích hợp thêm ứng dụng dịch vụ, loại bỏ ứng dụng dịch vụ cũ thì kết luận hệ thống có tính năng mở, có thể tích hợp được ứng dụng và có thể là giải pháp portal.

Khả năng xuất bản thông tin theo tiêu chuẩn (Content syndication):

Một trong những đặc tính quan trọng của portal là xuất bản thông tin cho người dùng cuối qua các tiêu chuẩn đã được công bố và thừa nhận trên toàn thế giới. Với các

dữ liệu được xuất bản theo tiêu chuẩn này, người dùng cuối có thể khai thác, sử dụng mà không cần thông qua giao diện tương tác của hệ thống mà sử dụng một số phần mềm của hãng thứ 3.Hiện tại có nhiều chuẩn xuất bản thông tin, nhưng tất cả các chuẩn xuất bản thông tin được ủng hộ và sử dụng nhiều nhất trên thế giới đều lấy cơ sở ngôn ngữ đánh dấu mở rộng XML (eXtensible Markup Language) làm nền tảng, đáng kể là RDF (Resource Description Format), RSS (Realy Simple Syndication), NITF (News Industry Text Format), NewsML và ATOM Syndication Format. Hiện tại có 2 tiêu chuẩn được sử dụng rộng rãi nhất là RSS và ATOM.

Để kiểm định tính năng này, bạn hãy yêu cầu nhà cung cấp trình diễn hoặc giới thiệu cách thức hệ thống xuất bản thông tin từ một hoặc nhiều module dịch vụ khác nhau thành các tài liệu theo tiêu chuẩn RSS hoặc ATOM. Tại đây có thể có nhiều kết quả khác nhau, như:

+ Nếu nhà cung cấp không có khái niệm gì về RSS hay ATOM, thì có thể kết luận ngay rằng hệ thống của nhà cung cấp này không có khả năng xuất bản thông tin theo tiêu chuẩn.

+ Nếu hệ thống có thể xuất bản tài liệu ra tiêu chuẩn RSS, nhưng cần phải "bẻ" mã chương trình ra chỉnh sửa lại thì có thể kết luận hệ thống có khả năng xuất bản thông tin với chuẩn nhưng không phải là portal.

+ Nếu có khả năng xuất bản ngay tức thì nội dung thành RSS, bạn hãy yêu cầu xuất bản thông tin có đầy đủ nội dung chứ không chỉ tóm tắt như tài liệu RSS đã cung cấp, nếu nhà cung cấp không thể làm được hoặc không thể đưa ra được hướng giải quyết cụ thể thì có thể kết luận rằng hệ thống có khả năng xuất bản thông tin theo tiêu chuẩn nhưng chưa đầy đủ.

+Nếu hệ thống cho phép xuất bản thành RSS và ATOM, chứa đầy đủ nội dung thông tin thì có thể kết luận hệ thống có khả năng đầy đủ để xuất bản thông tin với tiêu chuẩn công nghiệp.

+Nếu nhà cung cấp đưa ra được giải pháp đồng bộ dữ liệu giữa nhiều hệ thống bằng tài liệu theo tiêu chuẩn như ATOM hay SSE (Simple Sharing Extension for ATOM and RSS) thì có thể kết luận rằng đó là hệ thống rất mạnh trong xuất bản thông tin.

Hỗ trợ nhiều môi trường hiển thị thông tin (Multidevice support):

Đây là một tính năng phụ nhưng khá quan trọng vì với xu thế hiện tại, người sử dụng có thể dùng nhiều loại thiết bị để truy cập hệ thống tại nhiều địa điểm khác nhau.Để kiểm định tính năng này, bạn hãy yêu cầu nhà cung cấp trình diễn hoặc giới thiệu nội dung được hiển thị trên thiết bị cầm nay như PDA, Pocket PC, iPhone, Nokia 9500, ... Nếu không thể hiển thị được trên các thiết bị này, có thì kết luận là hệ thống không hỗ trợ hiển thị dữ lilệu ở môi trường và thiết bị khác nhau.

Khả năng đăng nhập một lần (Single Sign on - SSO):

Tính năng này là một trong các tính năng tối quan trọng của giải pháp portal, vì số lượng người dùng và dịch vụ ứng dụng sẽ tăng dần theo thời gian. Khi hệ thống cung cấp tính năng này, người sử dụng chỉ cần đăng nhập đúng một (01) lần duy nhất khi bắt đầu sử dụng hệ thống, mỗi khi dịch chuyển giữa các màn hình làm việc hoặc các module nghiệp vụ thì không cần phải đăng nhập lại, và khi đó các thành phần của hệ thống phải tự nhận biết được đó là người sử dụng nào, thẩm quyền đến đâu.Để kiểm định tính năng này, bạn hãy yêu cầu nhà cung cấp trình diễn hoặc giới thiệu cách thức đăng nhập hệ thống, sau đó sử dụng ít nhất là 3 module n ghiệp vụ (ví dụ: quản trị nội dung, diễn đàn, chia sẻ tài liệu). Tại đây có thể có nhiều kết quả khác nhau, như:

+Nếu mỗi khi dịch chuyển sang các module nghiệp vụ mới, người dùng phải đăng nhập lại thì kết luận hệ thống không hỗ trợ khả năng SSO, và đây không phải là giải pháp portal.

+ Nếu khi dịch chuyển giữa các module nghiệp vụ vẫn xác định được người dùng, bạn hãy đăng xuất (thoát - sign out/log out) và quay về sử dụng một module nghiệp vụ khác, nếu thấy hệ thống vẫn nhận ra người dùng (mặc dù đã sign-out) thì có thể kết luận đó là hệ thống giả lập tính năng SSO, và đó không phải là giải pháp portal.

+ Nếu đăng nhập và đăng xuất đều tốt (không bị lỗi trong 2 tình huống trên), thì có thể kết luận hệ thống có hỗ trợ SSO. Khi đó bạn hãy yêu cầu điều hướng sử dụng sang một tên miền khác đang dùng chính hệ thống này, nếu vẫn giữ được thông tin đăng nhập thì kết luận là đã hỗ trợ SSO tốt, nếu không thì kết luận là hỗ trợ SSO chưa tốt.

+ Đồng thời, bạn hãy yêu cầu nhà cung cấp kết nối với hệ thống quản trị người dùng chuyên nghiệp với tiêu chuẩn LDAP để xác thực người dùng (ví dụ: đăng nhập bằng tài khoản của Microsoft Windows Domain của chính doanh nghiệp bạn), nếu

không thể thực hiện thì kết luận rằng tính năng SSO chưa toàn vẹn, nếu được thì khẳng định tính năng SSO đã rất tốt.

Khả năng quản trị portal (Portal administration)

Tính năng này xác định cách thức hiển thị thông tin cho người dùng cuối với nhiều cách thức và nguồn khác nhau. Tính năng này không chỉ đơn giản là thiết lập các giao diện người dùng với các chi tiết đồ hoạ (look-and-feel), với tính năng này người quản trị phải định nghĩa được các thành phần thông tin, các kênh tương tác với người sử dụng cuối, định nghĩa nhóm người dùng cùng với các quyền truy cập và sử dụng thông tin khác nhau. Để kiểm định tính năng này, bạn hãy yêu cầu nhà cung cấp trình diễn hoặc giới thiệu cách thức điều chỉnh các màn hình hiển thị thông tin, tạo lập các nguồn thông tin khác nhau với nhiều thẩm quyền sử dụng thông tin.Tại đây có thể có nhiều kết quả khác nhau, như

+ Nếu nhà cung cấp phải "bẻ" mã (code) của hệ thống ra thì mới điều chỉnh hoặc bổ sung được các nguồn thông tin hay màn hình hiển thị thì có thể kết luận ngay hệ thống đó không phải là giải pháp portal.

+Nếu hệ thống cho phép điều chỉnh được, bạn hãy yêu cầu thay đổi các vị trí hiển thị của các khối thông tin, thay đổi các nội dung sẽ hiển thị trong một vài khối thông tin, nếu khi đó nhà cung cấp lại bắt buộc phải sửa mã chương trình thì kết luận ngay rằng hệ thống không có khả năng và đó không phải là giải pháp portal. Nếu được thì kết luận đó hệ thống có khả năng cho phép nhà quản trị thay đổi thông tin, nguồn tin, ... khi cần.

Khả năng quản trị người dùng (Portal user management)

Tính năng này cung cấp các khả năng quản trị người dùng cuối, tuỳ thuộc vào đối tượng sử dụng của hệ thống. Tại đây, người sử dụng có thể tự đăng ký trở thành thành viên hoặc được người quản trị tạo lập và gán quyền sử dụng tương ứng. Đồng thời, hệ thống phải hỗ trợ và tích hợp công việc quản trị và xác thực người dùng bằng tiêu chuẩn công nghiệp LDAP.

Mặt khác, phân quyền sử dụng phải mềm dẻo và có thể thay đổi được khi cần.Để kiểm định tính năng này, bạn hãy yêu cầu nhà cung cấp trình diễn hoặc giới thiệu cách thức đăng ký tài khoản hoặc người quản trị tạo lập tài khoản sử dụng mới trong hệ

thống, tạo lập các nhóm quyền sử dụng và gán các quyền sử dụng này cho thành viên. Tại đây có thể có nhiều kết quả khác nhau, như:

+Việc đăng ký tài khoản mới hoặc tạo lập tài khoản mới rất đơn giản, nhưng không thể tạo lập các nhóm quyền sử dụng mới mà chỉ dùng được các nhóm quyền sử dụng sẵn có của hệ thống, thì kết luận hệ thống không hỗ trợ khả năng quản trị người dùng, và đây không phải là giải pháp portal.

+Nếu việc đăng ký/tạo tài khoản mới và tạo lập các nhóm sử dụng mới suôn sẻ, hãy yêu cầu nhà cung cấp gán quyền sử dụng nào đó trong một module nghiệp vụ cụ thể với nhóm người sử dụng này. Sau khi thực hiện xong, người sử dụng mới không thể khai thác được theo quyền đã được cấp thì kết luận hệ thống không thực sự hỗ trợ quản trị người dùng vì đó chỉ là "giả lập", và khi đó hệ thống này không thể gọi là portal được. Nếu tất cả đều hoạt động tốt, kết luận là đã hỗ trợ tốt tính năng quản trị người dùng.

⇒ +Nếu hệ thống chỉ thoả mãn từ 5 tính năng nêu trên trở xuống (thoả mãn 5 hoặc thoả mãn ít hơn 5 tính năng) thì kết luận đó là ứng dụng web hoặc phần mềm quản trị nội dung chứ không phải là giải pháp portal.

+ Nếu thoả mãn 6 tính năng 1,2,3,5,6,7 mà không thoả mãn tính năng 4 (support multi-device) thì kết luận đó thực sự là giải pháp portal, và có ghi chú kèm bên cạnh là sử dụng tối ưu trên máy tính.

+Nếu thoả mãn tất cả cả 7 tính năng trên, thì đó thực sự là giải pháp portal và có khả năng hoạt động trên nhiều môi trường/thiết bị khác nhau

1.2.Các đặc trưng cơ bản của portal

1.2.1.Chức năng tìm kiếm (search function)

Chức năng tìm kiếm là dịch vụ đầu tiên cần phải có của tất cả các Portal. Sau khi người sử dụng mô tả loại thông tin mà mình cần thông qua các từ khoá hoặc tổ hợp các từ khoá, dịch vụ này sẽ tự động thực hiện tìm kiếm thông tin trên các Website có trên Internet và trả lại kết quả cho người dùng. Thời gian thực hiện của dịch vụ tìm kiếm này rất nhanh, do vậy rất tiện lợi cho người dùng.

1.2.2.Dịch vụ thư mục (Directory service)

Đối với những người dùng không muốn tìm kiếm thông tin qua các từ khoá, họ có nhu cầu tìm kiếm thông tin theo một chủ đề, lĩnh vực nào đó, thì có thể sử dụng dịch vụ thư mục phân loại thông tin. Dịch vụ thư mục là dịch vụ thực hiện phân loại và sắp xếp thông tin trên các website theo chủ đề có thể có nhiều chủ đề con trong một chủ đề và có thể tiếp tục phân tách xuống các mức thấp hơn.

1.2.3.Ứng dụng trực tuyến(Online desktop application)

Bao gồm các ứng dụng phổ biến nhất của Internet, hiện nay có các ứng dụng điển hình như :

- Thư điện tử: Các Portal lớn như Yahoo, Excite, v.v… thường cung cấp các tài khoản điện tử (E-mail account) miễn phí cho người dùng. Dịch vụ này rất có ý nghĩa vì người dùng có thể nhận/gửi tại bất cứ địa điểm nào của Internet.

- Lịch cá nhân: Một số Portal cung cấp dịch vụ "lịch cá nhân - calendar" miễn phí cho người dùng. Dịch vụ này giúp người sử dụng có thể sử dụng lịch cá nhân mọi nơi trên Internet.

- Hội thoại trực tuyến: Dịch vụ này cho phép nhóm người dùng hội thoại trực tuyến với nhau thông qua môi trường Internet, không phụ thuộc vào khoảng cách địa lý giữa họ. Có thể liệt kê nhiều loại dịch vụ trực tuyến khác như dịch vụ hỗ trợ kỹ thuật trực tuyến giữa các nhà sản xuất với khách hàng của mình…

- Các dịch vụ khác: Một trong những dịch vụ hấp dẫn người sử dụng là bưu thiếp điện tử. Thay vì gửi bưu thiếp qua đường bưu điện thông thường, ngay nay người sử dụng có thể gửi bưu thiếp chức mừng người thân của mình thông qua mạng Internet.

1.2.4.Cá nhân hoá dịch vụ (Personalization or Customization)

Cá nhân hoá là dịch vụ đặc trưng quan trọng của Portal. Trên cơ sở các thông tin của từng khách hàng cụ thể, nhà cung cấp có thể tạo ra các dịch vụ mang tính định hướng cá nhân, phù hợp với yêu cầu, sở thích của từng khách hàng riêng biệt của mình. Thông qua đó các nhà cung cấp có khả năng tăng cường mối quan hệ với khách hàng, duy trì được sự tín nhiệm của khách hàng đối với nhà cung cấp.

Cá nhân hoá các dịch vụ được tiến hành thông qua dữ liệu thông tin cá nhân về khách hàng (customer profiles). Dữ liệu này chứa các thông tin mang tính cá nhân như

nghề nghiệp, thói quen, sở thích v.v… từ những thông tin cá nhân này, các nhà cung cấp có khả năng giới hạn cung cấp các thông tin và các dịch vụ mà khách hàng thực sự quan tâm muốn có. Có nghĩa là tránh được việc cung cấp các thông tin và dịch vụ không cần thiết có thể sẽ gây khó chịu cho khách hàng, và thậm chí dẫn đến quyết định ngừng sử dụng dịch vụ của nhà cung cấp.

1.2.5.Cộng đồng ảo (Virtual community or Collaboration)

Cộng đồng ảo là một "một địa điểm ảo" trên Internet mà các cá nhân, các doanh nghiệp có thể "tập hợp" để giúp đỡ, hợp tác với nhau trong các hoạt động thương mại. Nói một cách khác "cộng đồng ảo" mang lại cơ hội hợp tác cho các cá nhân, tổ chức doanh nghiệp mà ranh giới địa lý không còn có ý nghĩa. Sau đây là một số ví dụ về cộng đồng ảo:

- Hội thoại trực tuyến – Online chat: Thông qua dịch vụ này người ta có thể triển khai các hội nghị mà không cần phải tập trung toàn bộ cán bộ công nhân viên ở các địa phương trong phạm vi cả nước về một địa điểm cụ thể nào đó.

- Hỗ trợ trực tuyến - Online support : Tại đây khách hàng có thể nhận được trực tiếp các hỗ trợ, tư vấn của các nhà sản xuất về sản phẩm mà khách hàng đã lựa chọn.

1.2.6.Một điểm tích hợp thông tin duy nhất (Comporate Portal)

Đặc trưng này cho phép đơn vị cung cấp cho người sử dụng dùng một điểm truy nhập duy nhất để thu thập và xử lý thông tin từ các nguồn khác nhau, hoặc sử dụng các ứng dụng để khai thác kho tài nguyên thông tin chung. Như chúng ta đã biết, có rất nhiều thông tin hàng ngày cần phải được xử lý và chuyển đến người dùng dưới nhiều nguồn khác nhau, ví dụ như E-mail, news, tài liệu, báo cáo, các bài báo, audio và các video files, v.v… sẽ rất khó khăn cho người dùng nếu các thông tin này được xử lý một cách riêng rẽ; Comporate Portal cho phép sử dụng các công cụ tích hợp để xử lý các nguồn thông tin này, do vậy năng suất lao động xử lý các thông tin của người dùng sẽ được nâng cao.

1.2..7.Kênh thông tin (Channel)

Portal cũng cho phép xây dựng các liên kết (connector) tới các ứng dụng hoặc Portal khác. Một Portal khác hoặc một Website thông thường khác có thể cung cấp nội dung thông tin của mình trong kênh thông tin của Portal. Kênh thông tin là đặc tính rất mới của Portal, cho phép xây dựng các dịch vụ truy cập, xử lý các thông tin nằm bên trong mạng Intranet của một tổ chức, và sau đó tổ chức hiển thị kết quả xử lý tin trên kênh thông tin của Portal.

1.3.Phân loại portal

Việc phân loại Portal có thể có nhiều cách khác nhau. Nếu căn cứ vào đặc trưng của Portal người ta chia Portal thành các loại như sau :

1.3.1.Consumer Portal

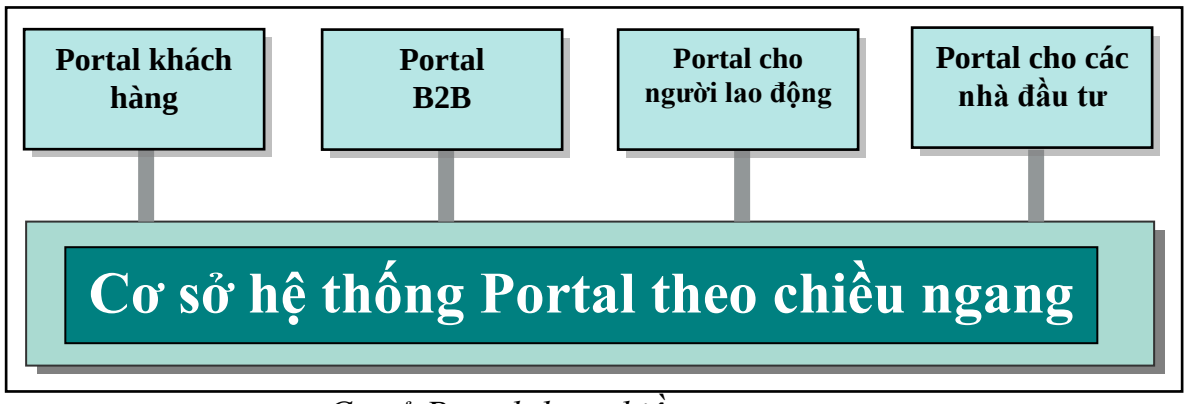
Cung cấp nhiều lựa chọn cho việc tìm kiếm, chuyển, E-mail, tự sửa khuôn dạng, lựa chọn tin tức, calendar, quản lý địa chỉ liên hệ, các cuộc hẹn, các lưu ý, chú thích, các địa chỉ website, real-time chat và các chức năng Intranet, v.v…

1.3.2.Vertical Portal

Chuyên cung cấp các thông tin và dịch vụ cho một lĩnh vực chuyên môn, khoa học, kinh tế cụ thể nào (mang tính chuyên ngành).

1.3.3.Horizontal Portal

Nội dung bao trùm nhiều chủ đề (mang tính diện rộng), phục vụ các mối quan tâm khác nhau, hỗ trợ bằng các chức năng dịch vụ phong phú, phục vụ cộng đồng, phục vụ tổ chức hành chính.



Cơ sở Portal theo chiều ngang

1.3.4.Enterprise Portal

Cung cấp các dịch vụ truy xuất thông tin từ mọi nguồn tài nguyên thông tin trong mạng Intranet của một tổ chức qua một cổng truy cập duy nhất.

1.3.5.B2B Portal

Cung cấp các dịch vụ định hướng theo mối quan hệ tương tác thông tin hai chiều giữa các doanh nghiệp (B2B) trong môi trường thương mại điện tử.

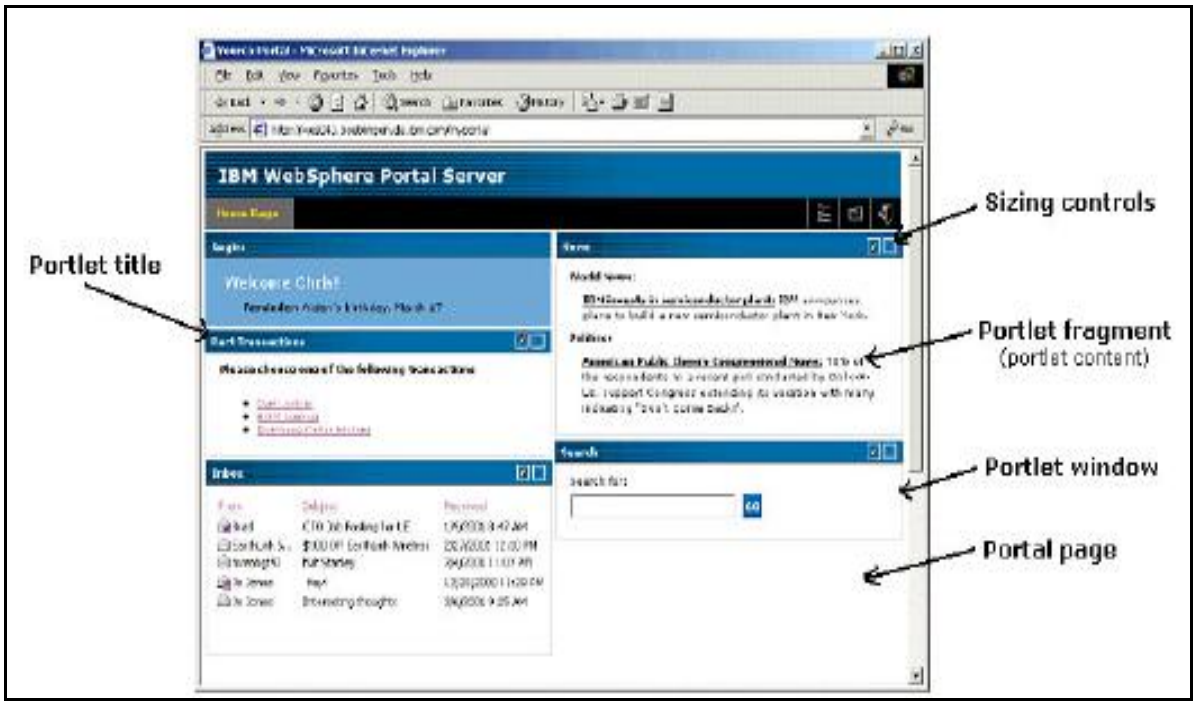
1.3.6.G2B Portal

Cung cấp các dịch vụ hành chính công theo mối quan hệ tương tác thông tin hai chiều giữa các cơ quan hành chính nhà nước (G2G) trong môi trường trao đổi thông tin điện tử.

1.4.Các kỹ thuật của hệ thống portal

1.4.1.Portlet

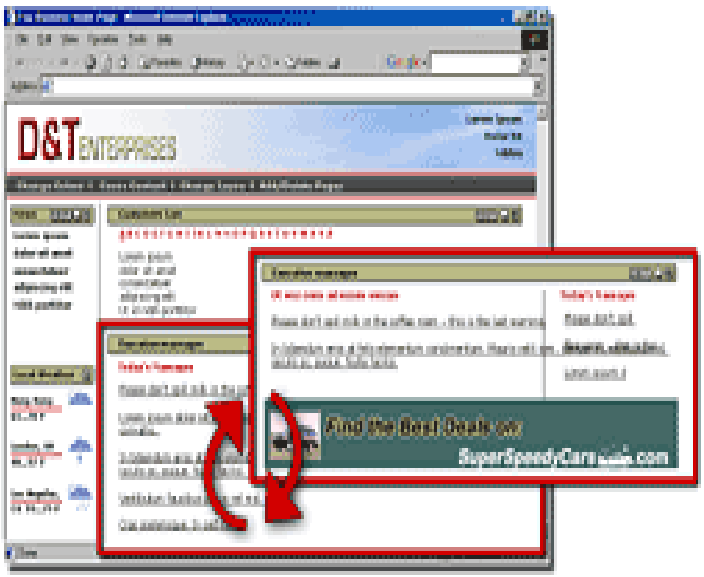
Portlet là giao diện người dùng, là các module tương tác nhiều mức cho phép tích hợp vào Portal các ứng dụng web khác nhau. Các Portlet này sinh ra các đoạn trang, các đoạn trang này được Portal ghép lại thành một trang hoàn chỉnh .



Các thành phần của một trang Portal

Portlet thực thi trong môi trường thời gian thực được gọi là Portlet Container, các Portlet trình bày nội dung của chúng trong một cửa sổ hiện trên trang Portal, tương tự

Nghiê



như cửa sổ trong màn hình (desktop). Cửa sổ của Portlet có một thanh tiêu đề chứa, các nút điều khiển cho phép người sử dụng mở rộng và thu nhỏ nó .

Một Portlet có thể hiển thị trên một trang web như một cửa sổ cá nhân nhỏ, Portlet là nội dung bên trong cửa sổ, nó không phải là bản thân cửa số đó.

Các Portlet bao gồm nhiều mức, cho phép người sử dụng giao tiếp với nó để thực hiện công việc trong môi trường Portal.

Các mức của Portlet có thể có trong Portal

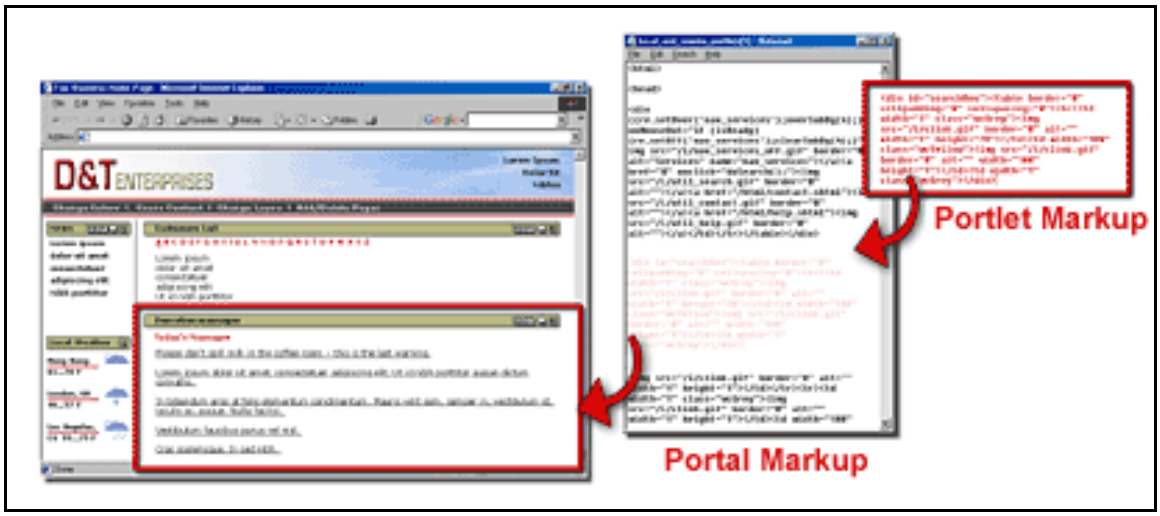
1.4.2.Phân loại portlet và các dịch vụ web

Giống như dịch vụ web hướng dữ liệu, các Portlet dựa trên kiến trúc hướng dịch vụ, nó cho phép các công ty sử dụng lại các thành phần của phần mềm để nhanh chóng xây dựng các ứng dụng trong các Portal mới.

Không giống như các dịch vụ web hướng dữ liệu, các Portlet tóm lược các dịch vụ tác nghiệp ở mức cao bao gồm các tương tác người dùng, các lưu đồ và các trình diễn tùy biến.

Portlet địa phương

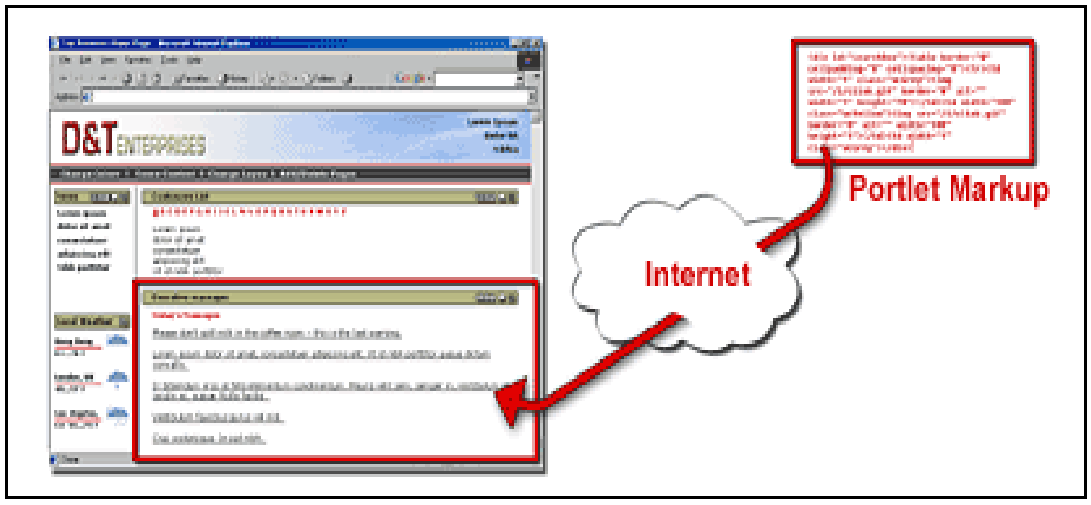
Các Portlet địa phương là các Portlet thực thi ở bên trong một máy chủ Portal. Khi một máy chủ Portal sinh ra một trang và những thứ cần thiết trong một đoạn trang, nó gọi Code Portlet và sử dụng giao diện tiền định nghĩa. JSR168 định nghĩa một giao diện Portlet địa phương chuẩn cho môi trường J2EE.



Các Portlet địa phương gọi tới Code Portlet

Portlet từ xa

Portlet từ xa là các Portlet thực thi bên ngoài một máy chủ Portal, hoặc bên trong một máy chủ của một tổ chức hoặc ở một vị trí từ xa. Khi một Portal cần đoạn trang, nó sẽ gọi Portlet từ xa thông qua SOAP.

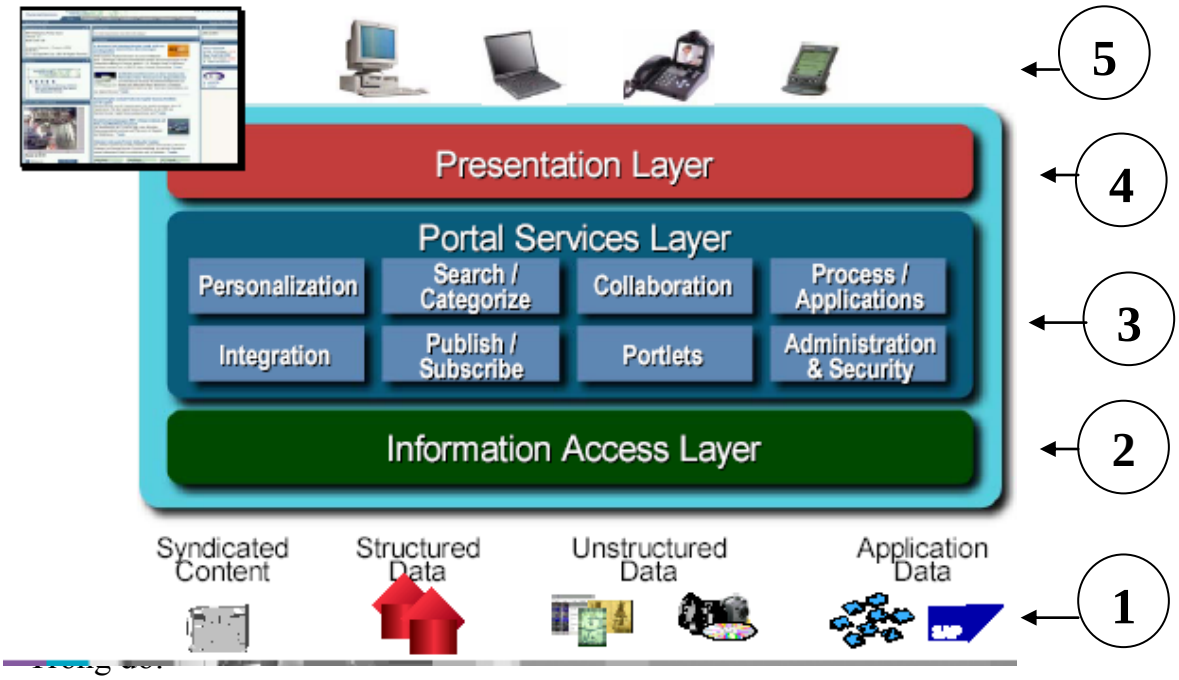


Trang Portal gọi đến từ một Portlet từ xa

Giao thức WSRP cung cấp định nghĩa một chuẩn giao diện SOAP cho các Portlet từ xa. Vấn đề quan trọng của Portlet từ xa là tách các Portlet ra khỏi tổ chức và môi trường Portal

1.5.Khung làm việc của hệ thống Portal

Hình ảnh về khung làm việc của hệ thống Portal được mô tả như sau:



1: Là các nguồn dữ liệu có cấu trúc, không cấu trúc, dữ liệu ứng dụng hoặc nội dung được cung cấp.

2: Tầng truy cập thông tin, làm nhiệm vụ truy cập tới mọi nguồn tài nguyên dữ liệu.

3: Tầng dịch vụ Portal, những dịch vụ đặc trưng tiêu biểu của Portal như: cá nhân hóa, tích hợp, dịch vụ tìm kiếm và phân loại, dịch vụ xuất bản và đặt báo, dịch vụ hợp tác, các ứng dụng, xử lý, quản trị và bảo mật.

4: Tầng trình diễn, ở đó Portal có nhiệm vụ tổng hợp thông tin thành một trang web và hiển thị theo yêu cầu của người dùng.

5: Các thiết bị truy cập mạng; Các thiết bị này truy cập Portal thông qua các kênh của Portal đó là các kênh dành cho mạng Intranet, mạng Internet, mạng không dây, v.v…

1.6.Các bước xây dựng Portal

1.6.1.Lập kế hoạch

Đây là giai đoạn xây dựng giải pháp tổng thể, đáp ứng nhu cầu quản lý và chiến lược của khách hàng. Kế hoạch tổng thể bao gồm: phạm vi của dự án, các mục tiêu

chiến lược của khách hàng và hiện trạng của hệ thống bao gồm cả các mối quan hệ thông tin nội bộ với bên ngoài.

1.6.2.Thiết kế tổng thể

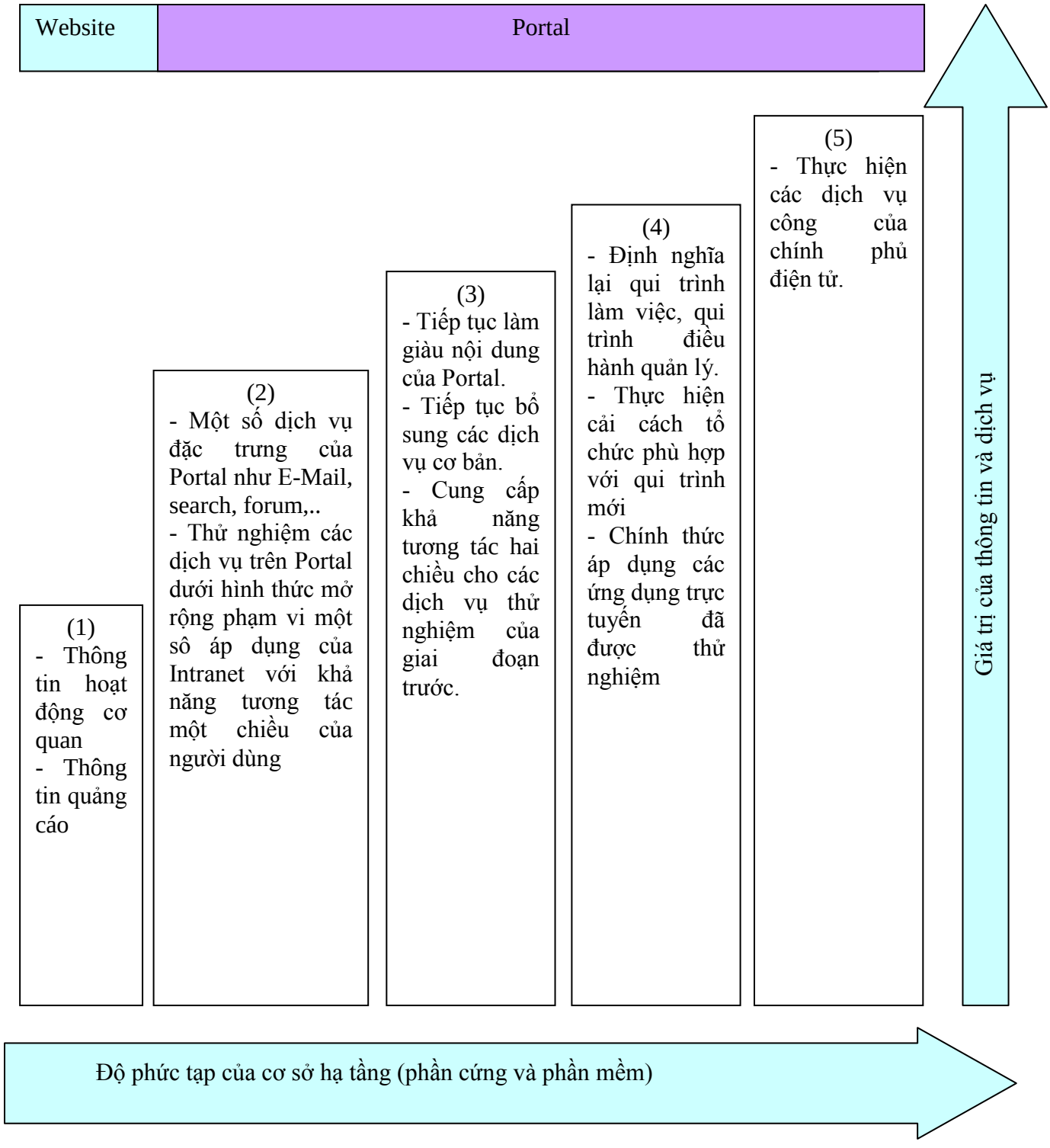
Thiết kế tổng thể là giai đoạn xây dựng kiến trúc ứng dụng cho phép chuyển hoá từ các yêu cầu nghiệp vụ sang ứng dụng Portal. Cũng như các phần mềm ứng dụng, kiến trúc ứng dụng bao gồm mô hình chức năng và mô hình hoạt động. Mô hình chức năng là toàn bộ các chức năng nghiệp vụ của hệ thống, mô tả cấu trúc, phân cấp các thành phần của hệ thống, các trao đổi thông tin và các giao diện giữa các thành phần của hệ thống. Mô hình hoạt động mô tả kiến trúc phần cứng (hạ tầng phần cứng, phương thức tổ chức mạng), kiến trúc phần mềm và các thành phần dữ liệu, các ràng buộc (tốc độ xử lý, mức độ bảo mật,…) và phần quản trị hệ thống (lập kế hoạch nguồn lực, chuyển giao hệ thống, sao lưu, khôi phục).

Kiến trúc ứng dụng cũng phải chỉ rõ mức độ đáp ứng của các giải pháp đối với chiến lược kinh doanh và phương thức đạt được yêu cầu đó.

1.6.3.Phát triển Portal

Phát triển là giai đoạn cài đặt giải pháp đã được xây dựng ở các bước trên, bao gồm: thiết kế, lập trình, kiểm tra, cài đặt sử dụng hệ thống Portal. Các phân tích viên thông thường tham gia vào giai đoạn này với vai trò kiểm soát viên để đảm bảo cho hệ thống đáp ứng được yêu cầu của người dùng.

Các giai đoạn hình thành và phát triển Portal được thể hiện qua sơ đồ sau



Các giai đoạn của lộ trình xây dựng và triển khai Portal

Chương 2

TỔ CHỨC DỮ LIỆU ,CƠ CHẾ CHUYỂN ĐỔI DỮ LIỆU TRONG CỔNG THÔNG TIN PHỤC VỤ CHO VIỆC TÌM KIẾM VÀ KHAI THÁC DỮ LIỆU

2.1.Tổ chức dữ liệu trông hệ thống thông tin

2.1.1.Một số mô hình tổ chức CSDL trong hệ thống Client \Server

Nhìn chung mọi ứng dụng CSDL đều bao gồm các phần: thành phần xử lý ứng dụng (Application processing components); thành phần phần mềm CSDL (Database software componets) và bản thân CSDL (The database itself) [4].

Các mô hình về xử lý CSDL khác nhau là bởi các trường hợp của 3 loại thành phần nói trên định vị ở đâu. Hiện nay, có các mô hình tổ chức CSDL của hệ thống Client/Server sau :

a. Mô hình CSDL tập trung (Centralized database model

Trong mô hình này, các thành phần xử lý ứng dụng, phần mềm CSDL và bản thân CSDL đều ở trên một bộ xử lý. Ví dụ người dùng máy tính cá nhân có thể chạy các chương trình ứng dụng có sử dụng phần mềm CSDL Oracle để truy nhập tới CSDL nằm trên đĩa cứng của máy tính cá nhân đó. Khi các thành phần ứng dụng, phần mềm CSDL và bản thân CSDL cùng nằm trên một máy tính thì ứng dụng đã thích hợp với mô hình tập trung. Hầu hết công việc xử lý luồng thông tin chính được thực hiện bởi nhiều tổ chức mà vẫn phù hợp với mô hình tập trung. Ví dụ một bộ xử lý mainframe chạy phần mềm CSDL IMS hoặc DB2 của IBM có thể cung cấp cho các trạm làm việc ở các vị trí phân tán sự truy nhập nhanh chóng tới CSDL trung tâm. Tuy nhiên, trong rất nhiều hệ thống như vậy, cả ba thành phần của ứng dụng CSDL đều thực hiện trên cùng một máy mainframe, do vậy, cấu hình này cũng thích hợp với mô hình tập trung.

b. Mô hình CSDL theo kiểu file - server (File - server database model)

Trong mô hình CSDL theo kiểu file - server các thành phần ứng dụng, phần mềm CSDL ở trên một hệ thống máy tính và các file vật lý tạo nên CSDL nằm trên hệ thống máy tính khác. Một cấu hình như vậy thường được dùng trong môi trường cục bộ, trong đó một hoặc nhiều hệ thống máy tính đóng vai trò của server, lưu trữ các file

dữ liệu cho hệ thống máy tính khác xâm nhập tới. Trong môi trường file server, phần mềm mạng được thi hành và làm cho các phần mềm ứng dụng cũng như phần mềm CSDL chạy trên hệ thống của người dùng đầu cuối, coi các file hoặc CSDL trên file server thực sự như là trên máy tính của chính họ. Mô hình file - server rất giống với mô hình tập trung. Các file CSDL nằm trên máy khác với các thành phần ứng dụng và phần mềm cơ sở dữ liệu; tuy nhiên các thành phần ứng dụng và phần mềm CSDL có thể có cùng thiết kế để vận hành một môi trường tập trung. Thực chất phần mềm mạng đã làm cho phần mềm ứng dụng và phần mềm CSDL tưởng rằng chúng đang truy nhập CSDL trong môi trường cục bộ. Một môi trường như vậy có thể phức tạp hơn mô hình tập trung bởi vì phần mềm mạng có thể phải thực hiện cơ chế đồng thời cho phép nhiều người dùng có thể truy nhập vào cùng cơ sở dữ liệu.

c. Mô hình xử lý từng phần CSDL (Database extract processing model)

Một mô hình khác trong đó một CSDL ở xa có thể được truy nhập bởi phần mềm CSDL, được gọi là xử lý dữ liệu từng phần. Với mô hình này, người sử dụng tại một máy tính cá nhân có thể kết nối với hệ thống máy tính ở xa nơi có dữ liệu mong muốn. Người sử dụng có thể tác động trực tiếp đến phần mềm chạy trên máy ở xa và tạo yêu cầu để lấy dữ liệu từ CSDL đó. Người sử dụng cũng có thể chuyển dữ liệu từ máy tính ở xa về chính máy tính của mình và có thể thực hiện việc sao chép bằng phần mềm CSDL trên máy cá nhân. Với cách tiếp cận này, người sử dụng phải biết chắc chắn là dữ liệu nằm ở đâu và làm như thế nào để truy nhập và lấy dữ liệu từ một máy tính ở xa. Phần mềm ứng dụng đi kèm cần phải có trên cả hai hệ thống máy tính để kiểm soát sự truy nhập dữ liệu và chuyển dữ liệu giữa hai hệ thống. Tuy nhiên, phần mềm CSDL chạy trên hai máy không cần biết rằng việc xử lý CSDL từ xa đang diễn ra vì người sử dụng tác động tới chúng một cách độc lập.

d.Mô hình CSDL Client/Server (Client/Server database model).

Trong mô hình CSDL Client/Server, CSDL nằm trên một máy khác với các máy có thành phần xử lý ứng dụng. Nhưng phần mềm CSDL được tách ra giữa hệ thống Client chạy các chương trình ứng dụng và hệ thống Server lưu trữ cơ sở dữ liệu. Trong mô hình này, các thành phần xử lý ứng dụng trên hệ thống Client đưa ra yêu cầu cho

phần mềm CSDL trên máy client, phần mềm này sẽ kết nối với phần mềm CSDL chạy trên Server. Phần mềm CSDL trên Server sẽ truy nhập vào CSDL và gửi trả kết quả cho máy Client. Mới nhìn, mô hình CSDL Client/Server có vẻ giống như mô hình file - server, tuy nhiên mô hình Client/Server có rất nhiều thuận lợi hơn mô hình file - server. Với mô hình file - server, thông tin gắn với sự truy nhập CSDL vật lý phải chạy trên toàn mạng. Một giao tác yêu cầu nhiều sự truy nhập dữ liệu có thể gây ra tắc nghẽn lưu lượng truyền trên mạng. Giả sử một người dùng đầu cuối tạo ra một truy vấn để lấy dữ liệu tổng số, yêu cầu đòi hỏi lấy dữ liệu từ 1000 bản ghi, với cách tiếp cận file - server nội dung của tất cả 1000 bản ghi phải đưa lên mạng, vì phần mềm CSDL chạy trên máy của người sử dụng phải truy nhập từng bản ghi để thoả mãn yêu cầu của người sử dụng. Với cách tiếp cận CSDL Client/Server, chỉ có lệnh truy vấn khởi động ban đầu và kết quả cuối cùng cần đưa lên mạng, phần mềm CSDL chạy trên máy lưu giữ CSDL sẽ truy nhập các bản ghi cần thiết, xử lý chúng và gọi các thủ tục cần thiết để đưa ra kết quả cuối cùng. Front-end software Trong mô hình CSDL Client/Server, thường nói đến các phần mềm front-end software và back-end software. Front-end software được chạy trên một máy tính cá nhân hoặc một workstation, đáp ứng các yêu cầu đơn lẻ riêng biệt, phần mềm này đóng vai trò của Client trong ứng dụng CSDL Client/Server và thực hiện các chức năng hướng tới nhu cầu của người dùng cuối cùng, phần mềm Front-end software thường được chia thành các loại sau:

- End user database software: Phần mềm CSDL này có thể được thực hiện bởi người sử dụng cuối trên chính hệ thống của họ để truy nhập các CSDL cục bộ nhỏ cũng như kết nối với các CSDL lớn hơn trên CSDL Server.
- Simple query and reporting software: Phần mềm này được thiết kế để cung cấp các công cụ dễ dùng hơn trong việc lấy dữ liệu từ CSDL và tạo các báo cáo đơn giản từ dữ liệu đã có.
- Data analysis software: Phần mềm này cung cấp các hàm về tìm kiếm, khôi phục, chúng có thể cung cấp các phân tích phức tạp cho người dùng.
- Application development tools: Các công cụ này cung cấp các khả năng về ngôn ngữ mà các nhân viên hệ thống thông tin chuyên nghiệp sử dụng để xây

dựng các ứng dụng CSDL của họ. Các công cụ ở đây bao gồm các công cụ về thông dịch, biên dịch đơn đến các công cụ CASE (Computer Aided Software Engineering), chúng tự động tất cả các bước trong quá trình phát triển ứng dụng và sinh ra chương trình cho các ứng dụng cơ sở dữ liệu.

- Database administration tools:Các công cụ này cho phép người quản trị CSDL sử dụng máy tính cá nhân hoặc trạm làm việc để thực hiện việc quản trị CSDL như định nghĩa các cơ sở dữ liệu, thực hiện lưu trữ hay phục hồi. Back-end software phần mềm này bao gồm phần mềm CSDL Client/Server và phần mềm mạng chạy trên máy đóng vai trò là Server cơ sở dữ liệu.

e. Mô hình CSDL phân tán (Distributed database model)

Cả hai mô hình File - Server và Client/Server đều giả định là dữ liệu nằm trên một bộ xử lý và chương trình ứng dụng truy nhập dữ liệu nằm trên một bộ xử lý khác, còn mô hình CSDL phân tán lại giả định bản thân CSDL có ở trên nhiều máy khác nhau.

2.1.2.Mô hình tổ chức dữ liệu trong portal

Trong Portal bao gồm các loại dữ liệu sau :

-Dữ liệu có cấu trúc: là dữ liệu được tổ chức (thường được phân cấp bởi các từ khóa) để dễ dàng tìm kiếm. Các dữ liệu có cấu trúc thường bao gồm các báo cáo, các phân tích, các truy vấn đã được lưu và các loại tin tức kinh tế, xã hội. Các dữ liệu cấu trúc này thường được phân tán rộng trong các server riêng lẻ trên Internet. Ví dụ, trong việc quản lý thông tin của ngành Giáo dục và Đào tạo, hồ sơ một giáo viên có thể được lưu tại một máy chủ nào đó, khi cần các máy chủ khác trong hệ thống thông tin có thể tự động lấy dữ liệu của giáo viên đó về.

-Dữ liệu phi cấu trúc: là nguồn dữ liệu không có tổ chức và nằm bên ngoài CSDL. Dữ liệu phi cấu trúc có thể là dữ liệu dạng text, âm thanh, hình ảnh hay đồ họa, chúng được trích ra từ các tài liệu văn phòng, E-mail, biên bản cuộc họp, và nhiều nguồn khác nhau...

- Như vậy:dữ liệu trong Portal rất nhiều loại lại được tổ chức theo mô hình CSDL phân tán trên các server ở các vị trí khác nhau trong hệ thống. Để

khai thác được các thông tin này thì việc chuyển đổi thông tin giữa các máy chủ cần phải có các cơ chế chuyển đổi thông tin nhất định.

2.2.Cơ chế chuyển đổi thông tin giữa các server trong portal

Như đã nói ở trên, thông tin trong Portal thường có cấu trúc rất khác nhau và được tích hợp từ nhiều nguồn khác nhau trong hệ thống máy chủ phân tán. Do vậy, để thiết lập được chuyển đổi thông tin trong hệ thống Portal, dữ liệu trong hệ thống cần được chuẩn hoá. Đây cũng chính là xu hướng trong quá trình trao đổi thông tin trong hệ thống phân tán.

Hiện nay, trên thế giới đã đưa ra nhiều các phương thức để chuẩn hoá thông tin trong quá trình trao đổi thông tin, trong số các công cụ chuẩn hoá thông tin thì công cụ XML và XSLT được sử dụng nhiều nhất. Vậy XML và XSLT là gì ? Tại sao chúng được sử dụng khá phổ biến ?.

XML (eXtensible Markup Language) được coi là một giải pháp chuẩn hoá thông tin dùng để trao đổi dữ liệu trong hệ thống tin trong hệ thống CSDL phân tán. Văn bản XML có cấu trúc dữ liệu đơn giản kiểu flat-text, có thể được xử lý bởi bất kỳ một trình soạn thảo ASCII thông thường nào và tác nhân con người hoàn toàn có thể đọc hiểu được nội dung của văn bản này một cách dễ dàng. Ưu điểm nổi bật của XML là:

- Tách phần dữ liệu ra khỏi sự thể hiện (presentation) của nó, ưu điểm này thể hiện khi có nhiều ứng dụng/thiết bị cùng hiển thị một văn bản XML. Ví dụ như khi truy nhập thông tin thể thao từ trình duyệt trên máy PC hay trên WAP mobile (chẳng hạn trên mobile chỉ cần thông tin hết sức ngắn gọn về tỉ số của trận bóng đá chứ không cần chi tiết màu mè, hình ảnh... như trên trình duyệt của PC)

- Trao đổi thông tin giữa các module khác nhau trong các hệ thống phân tán: XML được tạo ra với mục tiêu cung cấp một giải pháp chuẩn hoá cấu trúc dữ liệu trong việc trao đổi thông tin giữa các đối tác phần mềm khác nhau, mà không cần quan tâm bên nhận thông tin và quá trình xử lý thông tin sau đó. Với vai trò là bên nhận thông tin, văn bản XML thông thường được chuyển hoá thành dạng thức khác thích hợp hơn cho bên nhận trong quá trình xử lý thông tin tiếp theo

Công nghệ XSLT (eXtensible Stylesheet Language Transformations)

XSLT được phát triển bởi W3C, là một ngôn ngữ dùng để chuyển đổi dữ liệu có cấu trúc XML từ dạng mô hình này sang dạng mô hình khác, và thậm chí có cấu trúc hoàn toàn khác không phải là XML. **XSLT** là thành phần của một ngôn ngữ khác, đó là XSL (eXtensible Stylesheet Language). XSL được tạo ra để định dạng và thể hiện dữ liệu XML dưới nhiều dạng thức khác nhau, thành phần còn lại của XSL là XSL-FO (XSL Formatting Objects) có nhiệm vụ làm nốt chức năng định dạng dữ liệu trong văn bản XML.

Với XSLT cấu trúc dữ liệu nguồn là XML, nhưng cấu trúc đích thì không nhất thiết phải là XML, có thể là HTML như trong thí dụ chuyển đổi từ XML sang HTML để hiển thị nội dung của văn bản XML lên trình duyệt. Chuyển đổi dữ liệu từ XML sang XML có mô hình dữ liệu khác được ứng dụng ở mức thấp trong các hệ thống thương mại điện tử phân tán.
XSLT được tạo ra dưới dạng một văn bản flat-text đơn thuần, văn bản này được gọi là stylesheet, mỗi stylesheet bao gồm nhiều template (được coi như là các function của XSL stylesheet). Xpath là một ngôn ngữ độc lập nhưng nó lại được ứng dụng rất nhiều trong các XSL stylesheet và nó được coi như là một ngôn ngữ con của XSLT. Nếu cấu trúc dữ liệu nguồn không phải là XML thì nó phải được định dạng lại thành cấu trúc XML trước khi sử dụng XSLT. Có nhiều thư viện sẵn có để làm việc này, như định dạng (convert) HTML thành XML hay thậm chí cho phép định dạng một số cấu trúc dữ liệu cũ để lại.

Đặc điểm cơ bản của XSLT

- Cú pháp của XSL/XSLT tuân theo cú pháp XML.

- Không gây ảnh hưởng phụ: Đây là một tính chất của các ngôn ngữ lập trình và ít được nhắc đến vì hầu hết các ngôn ngữ lập trình thông thường đều có tính side-effect. Các hàm (template) của XSLT lại không có tính chất side-effect, có nghĩa là không làm thay đổi giá trị các biến trong stylesheet, kết quả trả về của chúng luôn cố định và không phụ thuộc vào số lần được gọi hay thứ tự được gọi.

- Template dựa trên luật: XSLT stylesheet bao gồm một tập hợp các template, mỗi một template sử dụng luật để chỉ ra các thành phần dữ liệu XML (element) cụ thể sẽ được xử lý trong template đó, các luật ở đây sử dụng biểu thức Xpath. Như vậy, mỗi một node trong văn bản XML thường phù hợp với tiêu chí xử lý của một template nào đó trong stylesheet.

- Kết quả chuyển đổi không phụ thuộc vào ngôn ngữ lập trình: XSLT là một chuẩn công nghệ, các nhà cung cấp muốn sản phẩm của mình hỗ trợ XSLT thì họ phải tuân theo đặc tả công nghệ của XSLT. Kết quả của quá trình chuyển đổi hoàn toàn không phụ thuộc vào ngôn ngữ lập trình cũng như vài trò của các nhà cung cấp, mặc dù mỗi nhà cung cấp có thể đưa ra một thư viện, được gọi là XSLT transformer, có cách thức xử lý và chuyển đổi hoàn toàn khác nhau cũng như mức độ hỗ trợ công nghệ này trong thư viện của họ.

- Ngôn ngữ XSLT : XSLT là một ngôn ngữ vì thế nó cũng có một bộ lệnh riêng như một ngôn ngữ lập trình thông thường, ví dụ như lệnh lặp, rẽ nhánh, gọi hàm bên ngoài, truyền tham số... Nó cũng có các biến với các kiểu cơ bản như string, numeric, boolean... hoặc các biến có kiểu là XML element/node và các hàm thao tác trên chúng.

Các template trong XSL stylesheet được nằm trong một node gốc là "xsl:stylesheet? node này có các thuộc tính mô tả thông tin của stylesheet hiện thời như xsl version, xsl transformer và xsl formating object [18].

Một ví dụ về việc chuẩn hoá thông tin trong mô hình phân tán sử dụng XML đó là việc ra đời chuẩn MARC.

Vậy MARC là gì ? MARC (MAchine Readable Cataloging - Danh mục máy đọc được) là một hệ thống được phát triển bởi thư viện Quốc hội Hoa Kỳ vào năm 1966, để các thư viện có thể chia sẻ những dữ liệu thư mục máy đọc được (Machine-Readable Bibliographic Data). Có nghĩa là các hệ thống quản trị thư viện tự động phải cần phải có một dạng thức chung để có thể trao đổi dữ liệu với nhau. Hiện nay MARC21 sử dụng XML đang trở thành chuẩn phổ biến để các tổ chức, quốc gia trên thế giới áp dụng khi xây dựng hệ thống thư viện điện tử của mình.

Để có thể trao đổi thông tin trong hệ thống CSDL phân tán chúng ta cần phải xây dựng được mô hình khai thác thông tin.

2.3.Mô hình khai thác và tìm kiếm thông tin trong hệ thông tin

Mô hình xử lý CSDL trong hệ thống thông tin phân tán bao gồm: Master/Slave, mô hình Client/Server hay mô hình Server/Server .

2.3.1.Mô hình xử lí Mater/slave

Trong mô hình này, một hệ thống máy được gọi là slave thực hiện các công việc của chỉ thị bởi hệ thống master. Như vậy, các ứng dụng chạy trên môi trường Master/Slave dường như có tính phân tán, mặc dù việc phân tán xử lý này có một chiều từ Master đến Slave

2.3.2.Mô hình Client/Server

Hiện nay mô hình này được sử dụng rộng rãi trong môi trường CSDL phân tán, là mô hình xử lý giữa client và server. Các yêu cầu của client được gửi lên server, server xử lý các yêu cầu này rồi trả lại kết quả cho client.

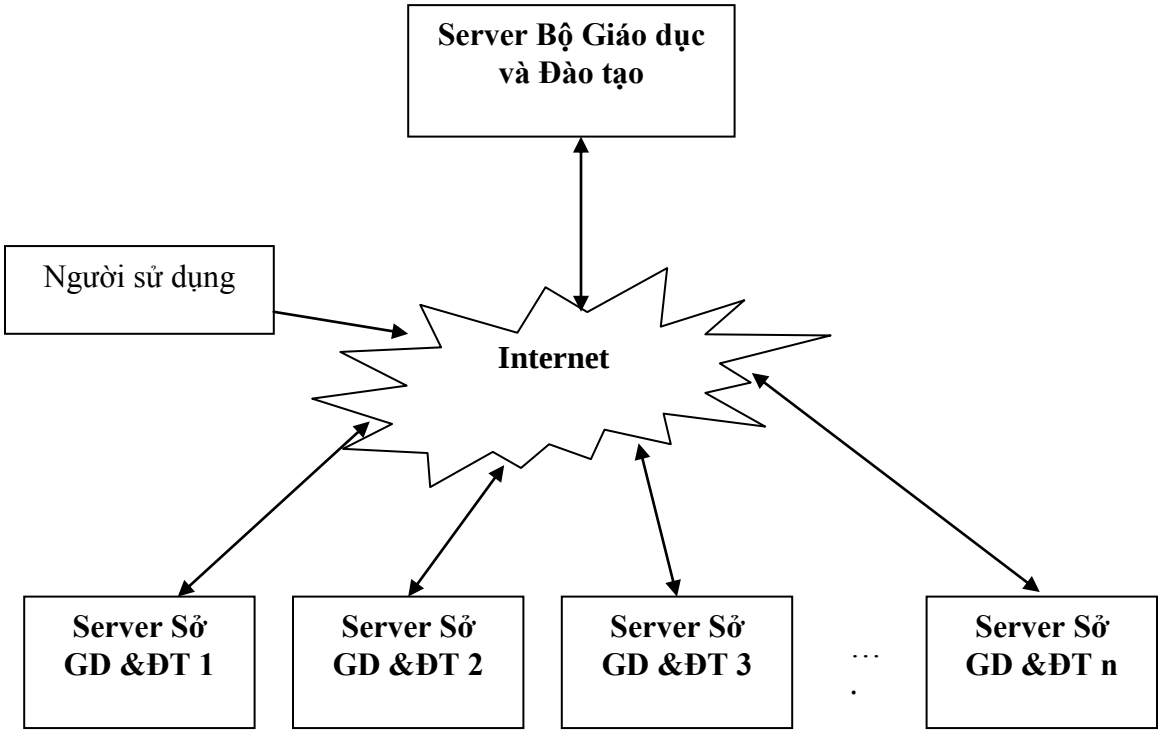
Mô hình client/server là mô hình ở mức cao hơn so với việc xử lý chia xẻ thiết bị thường thấy ở mạng LAN. Ví dụ, nếu một ứng dụng chạy trên một PC cần một bản ghi từ một tệp được chia sẻ nào đó, nó gán yêu cầu đọc toàn bộ tệp đó từ file server, sau đó ứng dụng phải tìm bản ghi đó trên tệp nhận được. Tài nguyên của file server được sử dụng để truyền cả tệp đó, trong khi tài nguyên của PC phải chia sẻ cho một chương trình tìm kiếm bản ghi trên tệp. Điều đó dẫn tới tài nguyên được sử dụng không hiệu quả và có thể dẫn tới quá tải trên đường truyền. Trong trường hợp Server CSDL ứng dụng chạy trên PC gửi yêu cầu đọc một bản ghi cho ứng dụng yêu cầu, như vậy cả client và server cùng hợp tác để thực hiện việc truy xuất dữ liệu .

2.3.3.Mô hình xử lí Server/Server

Là mô hình được sử dụng trong quá trình tích hợp thông tin trong cổng thông tin điện tử Portal, ở đây CSDL được phân tán trên nhiều server. Để có thể khai thác được hiệu quả thông tin nhằm hỗ trợ cho người sử dụng cần có một mô hình trao đổi thông tin một cách tự động giữa các Database Server. Ví dụ, trong ngành quản lý giáo dục của

Bộ Giáo dục và Đào tạo, tại các server của Sở Giáo dục và Đào tạo có đầy đủ thông tin về các trường THPT do đơn vị mình quản lý. Nếu người sử dụng muốn tìm hiểu về thông tin của trường THPT Mỹ Đức A, thì server của Bộ Giáo dục và Đào tạo sẽ gửi yêu cầu của người sử dụng đến tất cả các máy chủ của các Sở, sau quá trình trao đổi giữa các server sẽ trả lại người sử dụng thông tin mà người sử dụng yêu cầu.

Có thể mô hình hoá việc kết nối giữa các server trong cổng thông tin giáo dục bằng sơ đồ sau đây :



Mô hình Server/Server trong khai thác thông tin

Về mặt kiến trúc, mô hình xử lý Server/Server có các yêu cầu sau:

- Truyền thông phải tin cậy giữa các server.
- Phải có cơ chế điều khiển tránh tắc nghẽn giữa các server khi có khối lượng lớn thông tin được chuyển về máy yêu cầu cùng một lúc.
- Tại các server phải được cài đặt các module truy vấn, khi có yêu cầu truy vấn sẽ tự động thực hiện các yêu cầu và gửi lại kết quả cho máy yêu cầu

- Server yêu cầu cần phải có sự quản lý các kết quả gửi về từ các server khác trên mạng.

Để giải quyết được vấn đề trên chúng ta cần phải có các giải pháp khắc phục một số yêu cầu trong khi xây dựng mô hình này.

- Để đảm bảo quá trình tìm kiếm được thông suốt cần có cơ chế kiểm tra cơ chế Online của các server trong hệ thống cần khai thác thông tin, tránh tình trạng quá trình tìm kiếm bị dừng khi một trong các server trong hệ thống không Online.

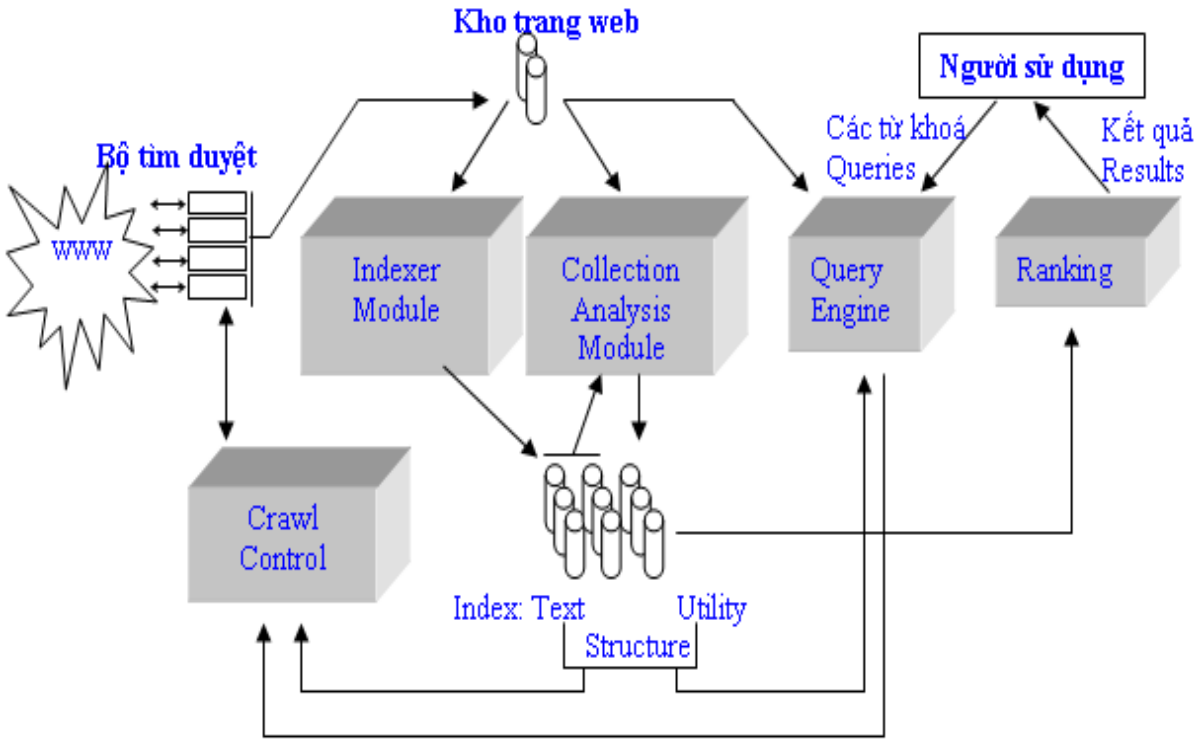
- Để tránh tình trạng tắc nghẽn đường truyền khi số lượng quá lớn kết quả gửi về máy trung tâm, chúng ta cần phải kiểm soát số lượng kết quả nhận được của từng server. Từ đó, có các biện pháp để hạn chế số lượng kết quả về trong cùng một khoảng thời gian bằng cách cắt nhỏ kết quả trong quá trình gửi về server trung tâm.

2.4.Một số thuật toán tìm kiếm dữ liệu trong hệ thống thông tin phân tán

2.4.1.Cấu trúc cơ bản của máy tìm kiếm

Máy tìm kiếm bao gồm các thành phần cơ bản sau đây:

Bộ tìm duyệt (Crawler): Hầu hết các máy tìm kiếm hoạt động đều dựa vào các bộ tìm duyệt. Bộ tìm duyệt là một chương trình nhỏ đảm nhận chức năng cung cấp dữ liệu (các trang web) cho máy tìm kiếm hoạt động. Bộ tìm duyệt thực hiện công việc duyệt web và tìm các mối liên hệ giữa các trang web này với các trang web khác. Các bộ tìm duyệt được cung cấp địa chỉ URL xuất phát, đọc trang web tương ứng, phân tích và tìm ra các URL có trong trang web đó, sau đó bộ tìm duyệt cung cấp các URL kết quả cho bộ điều khiển tìm duyệt (Crawl control). Bộ điều khiển tìm duyệt sẽ quyết định xem URL nào sẽ được duyệt tiếp theo và gửi kết quả về quyết định cho bộ tìm duyệt, bộ tìm duyệt cũng chuyển luôn các trang web đã duyệt vào kho trang web (Page Repository), các bộ tìm duyệt tiếp tục đi thăm các trang web khác trên Internet cho đến khi các nguồn chứa cạn kiệt.



Mô hình cấu trúc máy tìm kiếm

Bộ tạo chỉ mục (Indexer module) thực hiện việc khảo sát tất cả các từ khoá trong từng trang web có trong kho trang web, ghi lại các địa chỉ URL của các trang web có chứa mỗi từ. Kết quả sinh ra một bảng chỉ mục lớn. Nhờ có bảng chỉ mục này, máy tìm kiếm cung cấp tất cả các địa chỉ URL của các trang web khi có yêu cầu, khi cho một từ khoá bất kỳ qua bảng chỉ mục, máy tìm kiếm sẽ nhận được tất cả các URL của các trang web có chứa từ khoá đó. Chỉ mục này được gọi là chỉ mục nội dung. Việc tạo chỉ mục cho một hệ thống web thực sự là một việc làm rất khó khăn do kích thước đồ sộ của hệ thống web.

Bộ phân tích tập (Collection analysis module) hoạt động dựa vào các thuộc tính của bộ truy vấn (Query Engine). Ví dụ nếu bộ truy vấn đòi hỏi việc tìm kiếm hạn chế trong một số website đặc biệt thì công việc sẽ nhanh và hiệu quả hơn khi phải xây dựng một bảng chỉ mục các website mà trong đó có kết nối mỗi tên miền tới một danh sách

các trang web thuộc miền đó. Công việc như thế được thực hiện bởi bộ phân tích tập, nó sử dụng thông tin từ hai loại chỉ mục cơ bản (chỉ mục nội dung và chỉ mục cấu trúc) do bộ tạo chỉ mục cung cấp cùng với thông tin từ khoá trang web, các thông tin được sử dụng bởi phương pháp tính hạng (ranking) để tạo ra các chỉ mục tiện ích.

Bộ truy vấn (**Query Engine)** chịu trách nhiệm nhận các yêu cầu của người sử dụng. Bộ phận này hoạt động thường xuyên dựa vào bảng chỉ mục và thỉnh thoảng dựa vào kho trang web. Do số lượng các trang web là rất lớn, mà trong thực tế thì người sử dụng chỉ đưa vào một hoặc vài từ khoá, cho nên tập kết quả thường rất lớn, vì thế bộ xếp hạng có chức năng sắp xếp kết quả thành một danh sách các trang web theo thứ tự giảm dần về độ liên quan tới vấn đề mà người sử dụng đang quan tâm, và sau đó hiển thị danh sách kết quả cho người dùng.

2.4.2.Phương pháp biểu dữ liệu trong máy tìm kiếm

Biểu diễn chỉ mục nội dung

Chỉ mục nội dung trợ giúp việc tìm kiếm theo nội dung, giúp cho máy tìm kiếm có thể sử dụng bất cứ một phương pháp truy nhập truyền thống nào để tìm kiếm trong bộ dữ liệu. Máy tìm kiếm sử dụng chỉ mục liên kết ngược cho việc biểu diễn tài liệu.

Biểu diễn chỉ mục cấu trúc

Trong quá trình tạo chỉ mục, bộ tạo chỉ mục sẽ phân tích tất cả các siêu liên kết có trong tất cả các trang web và lưu trữ mọi thông tin quan trọng về các siêu liên kết đó trong file neo (anchor file). Các file này chứa đầy đủ các thông tin để xác định mỗi siêu liên kết xuất phát từ đâu và đi đến đâu cũng như cụm từ được dùng để đặt cho siêu liên kết. Một chương trình con của bộ tạo chỉ mục có chức năng chuyển địa chỉ quan hệ giữa các siêu liên kết thành địa chỉ tuyệt đối, và đưa địa chỉ đó vào thành phần trang web (docID), đồng thời sinh ra CSDL các siêu liên kết, trong đó có chứa từng đôi định danh trang web tương ứng với mỗi siêu liên kết. CSDL siêu liên kết dùng để tính hạng cho tài liệu.

2.4.3.Hoạt động của máy tìm kiếm Google

Thuật ngữ “Cỗ máy tìm kiếm ” được dùng chung để chỉ 2 hệ thống tìm kiếm: Một do các chương trình máy tính tự động tạo ra (Crawler-Based Search

Engines) và dạng thư mục internet do con người quản lý (Human-Powered Directories).Hai hệ thống tìm kiếm này tìm và lập danh mục website theo 2 cách khác nhau.

a. Crawler-Based Search Engines - Hệ thống tìm kiếm trên nền tự động

Những cỗ máy tìm kiếm tự động, như Google, tạo ra những danh sách của họ tự động. Chúng sử dụng các chương trình máy tính, được gọi là "robots", "spiders", hay crawlers để lần tìm thông tin trên mạng. khi có ai đó tìm kiếm một thông tin, các Search Engine lập tức hiển thị các thông tin lưu trữ tương ứng. Nếu bạn thay đổi những trang web của các bạn, những cỗ máy tìm kiếm tự động dần dần tìm thấy những sự thay đổi này, và điều đó có thể ảnh hưởng đến bạn được liệt kê như thế nào. Những tiêu đề trang, nội dung văn bản và các phần tử khác đều giữ một vai trò nhất định.

b. Human-Powered Directories - Các thư mục do con người quản lý và cập nhật

Các thư mục Internet - ví dụ như Dự án thư mục mở - Open Directory Project (Dmoz.org) hòan tòan phụ thuộc vào sự quản lý của con người. Bạn đăng ký website của bạn vào thư mục với một vài dòng mô tả ngắn gọn hoặc các biên tập viên của thư mục viết giúp phần mô tả cho bạn - chúng phù hợp với nội dung và chủ đề của từng danh mục.

Việc thay đổi những trang web của các bạn không có hiệu lực trên danh mục của các bạn. Những thứ hữu ích để cải thiện vị trí xếp hạng với một cỗ máy tìm kiếm không có gì để làm với việc cải thiện một vị trí trong một thư mục. Ngoại lệ duy nhất là một site tốt, với nội dung tốt, có lẽ thích hợp hơn để được xem xét so với một website nghèo nàn.

c."Hybrid Search Engines" - Các hệ thống tìm kiếm tổng hợp

Ngày trước, mỗi cỗ máy tìm kiếm sử dụng giải thuật riêng để tạo sự khác biệt. Đã là hệ thống tìm kiếm tự động thì không kèm theo một thư mục internet và

ngược lại. Nhưng hiện nay, hầu hết hệ thống tìm kiếm đều là sự tổng hợp của hệ thống tìm kiếm tự động và một thư mục do con người quản lý. Ví dụ, Yahoo có Yahoo Directory, Google có Google directory (dựa trên thư mục Dmoz), MSN và các hệ thống tìm kiếm khác cũng vậy.

d.Các thành phần của một cỗ máy tìm kiếm tự động

Những cỗ máy tìm kiếm tự động có ba phần tử chính. Đầu tiên là spider, cũng được gọi là crawlers. Spider đến thăm một trang web, đọc nó, và sau đó đi theo sau những mối liên kết tới những trang khác bên trong website. Có nghĩa là, khi có ai đó tìm kiếm đến một trang, các spiders sẽ ghi nhớ điều đó. Nó sẽ quay lại trang đó và theo chu kỳ 1-2 tháng. Như vậy, nếu trang web được tìm thấy càng nhiều, thì các spiders càng năng quay trở lại hơn và như thế, kết quả tìm kiếm của bạn cũng được cải thiện theo.

Mọi thứ spider tìm thấy đi vào trong phần thứ hai của cỗ máy tìm kiếm, Chỉ mục (the index). Chỉ mục, đôi khi gọi là tài liệu, là một kho lưu trữ khổng lồ chứa đựng một sự sao chép của mọi trang web mà spider tìm thấy. Nếu một trang web thay đổi, thì danh sách này được cập nhật với thông tin mới.

Đôi khi, cần phải có thời gian để các spiders lập chỉ mục cho một trang mới hay một trang được thay đổi nội dung. Như vậy, sẽ có trường hợp: một trang đã được các spiders tìm đến, nhưng lại chưa được lập chỉ mục. Và trong khỏang thời gian này, trang web sẽ hòan tòan không tồn tại trên Search engine.

Phần mềm tìm kiếm chính là phần tử thứ ba của một cỗ máy tìm kiếm. Đây là một chương trình máy tính có chức năng sàng lọc thông tin từ hàng triệu trang tương tự nhau để sắp xếp vị trí từng trang sao cho phù hợp nhất. Đây chính là nơi mà các công ty SEO khai thác để đưa một website nào đó lên vị trí Top khi được tìm kiếm với một hay nhiều từ khóa chỉ định.

2.5.Mô hình tìm kiếm thông tin trong CSDL phân tán

Việc tìm kiếm được thực hiện qua các bước sau:

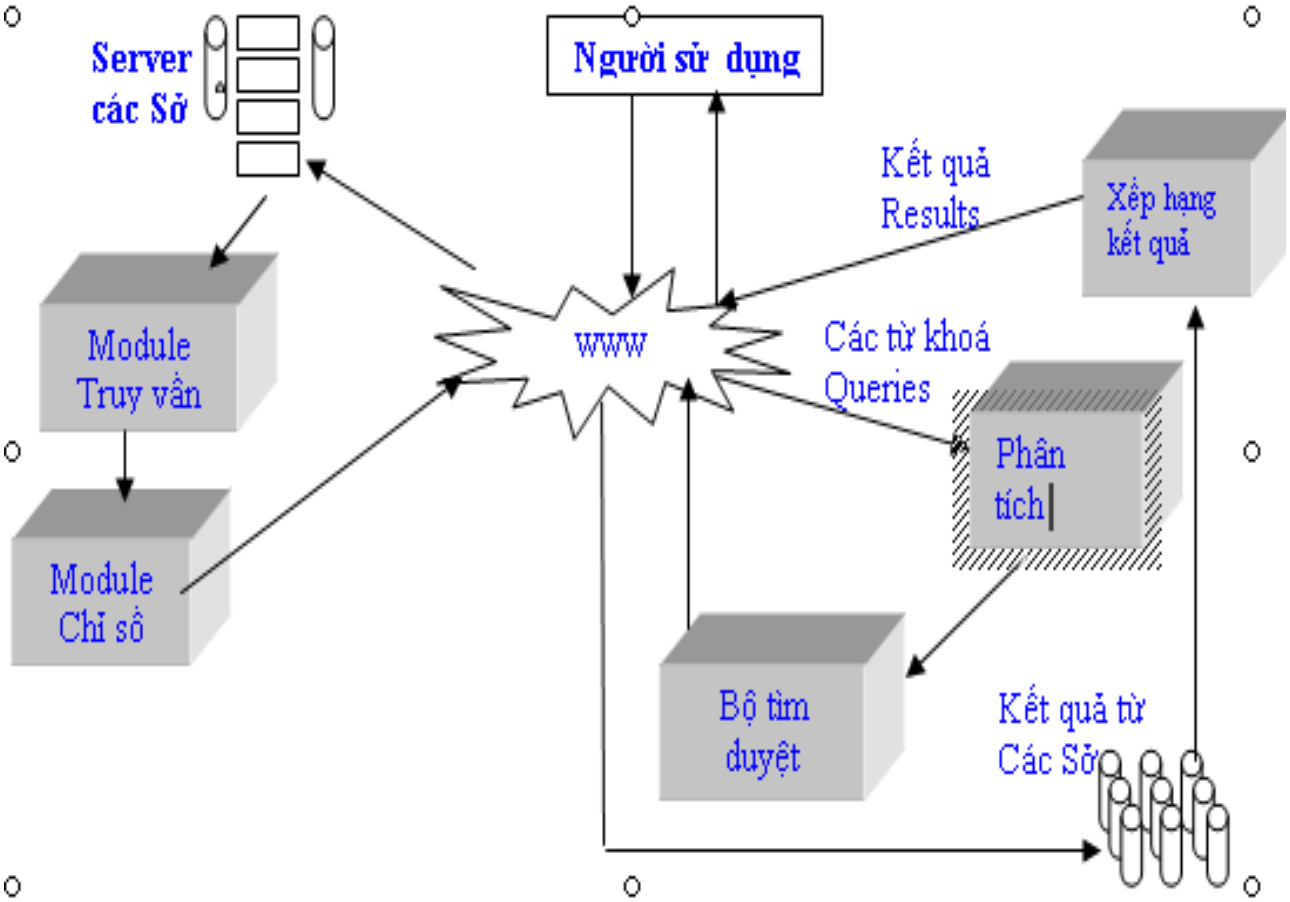
B1. Phân tích các yêu cầu của người sử dụng chuyển thành dạng biểu thức của ngôn ngữ SQL.

B2. Lần lượt gửi truy vấn này đến tất cả các máy chủ có kết nối đến máy chủ hiện tại.

B3. Tại các máy chủ khi nhận được yêu cầu từ máy chủ hiện tại, Module truy vấn tự động thực hiện các yêu cầu và gửi kết quả tìm kiếm về máy chủ yêu cầu.

B4. Tại máy chủ yêu cầu, kết quả sẽ được xếp hạng trước khi trả kết quả cho người sử dụng.

Có thể tóm tắt mô hình khai thác và tìm kiếm thông tin trong hệ thống thông tin phân tán qua sơ đồ sau :



Mô hình tìm kiếm trong hệ thống thông tin giáo dục

Hoạt động của mô hình khai thác và tìm kiếm thông tin được diễn giải như sau :

Người sử dụng thông qua Internet gửi các yêu cầu tìm kiếm tới máy chủ trung tâm. Tại đây bộ phân tích Query sẽ chuyển các yêu cầu của người sử dụng thành biểu thức dạng SQL. Từ đây dữ liệu sẽ chuyển đến bộ tìm duyệt. Bộ tìm duyệt thực hiện các kết nối đến các server của các Sở Giáo dục và Đào tạo thông qua môi trường Internet.

Tại server của các Sở Giáo dục và Đào tạo, khi nhận được yêu cầu module truy vấn sẽ thực hiện các yêu cầu từ server Bộ Giáo dục và Đào tạo. Kết quả sẽ được chuyển sang module chỉ số hoá, và được chuyển về server Bộ Giáo dục và Đào tạo. Tại server Bộ Giáo dục và Đào tạo sẽ tập kết quả của các server của các sở Giáo dục và Đào tạo, kết quả này được chuyển sang bộ xếp hạng kết quả trước khi gửi kết quả cho người sử dụng.

Chương 3

ÁP DỤNG NGHIÊN CỨU BÀI TOÁN GIẢI QUYẾT VẤN ĐỀ KHAI THÁC VÀ TÌM KIẾM THÔNG TIN TRONG CỔNG THÔNG TIN CỦA NGÀNH GIÁO DỤC VÀ ĐÀO TẠO

3.1.Yêu cầu khai thác, tìm kiếm thông tin của ngành Giáo dục.

3.1.1.Yêu cầu khai thác thông tin từ cơ sở:

Nhờ có chương trình hỗ trợ phát triển CNTT đưa tin học vào nhà trường của Chính phủ, mà phần lớn các phòng Giáo dục và Đào tạo, trường THPT, THCS trong phạm vi cả nước đã được trang bị máy tính và được cấp một account kết nối với Internet để phục vụ cho công tác quản lý, công tác dạy và học trong các nhà trường. Thực tế cho thấy kết quả của dự án này vẫn chỉ dừng lại ở công tác văn phòng, và dạy nghề phổ thông đối với các trường được trang bị số lượng lớn máy tính. Nguyên nhân của tình trạng trên là do : Cơ sở hạ tầng viễn thông vẫn còn hạn chế đối với các vùng miền núi, vùng sâu, vùng xa ; Nguồn nhân lực con người làm việc trong lĩnh vực này ngành Giáo dục và Đào tạo còn thiếu và yếu … Do vậy, công tác quản lý, tìm kiếm, lập báo cáo giáo dục nói chung vẫn chỉ dừng ở mức vừa bằng tay vừa bằng máy, máy tính vẫn chỉ là công cụ thay chiếc máy đánh chữ. Chính vì vậy, công tác quản lý và lập báo cáo còn mất nhiều thời gian, số liệu thì chưa thật chính xác. Để khắc phục tình trạng trên có thể đưa ra các nhu cầu cụ thể cần thiết áp dụng CNTT trong quá trình quản lý Giáo dục và Đào tạo của các cấp cơ sở như sau:

1. Đối với các trường THPT :

- Quản lý hồ sơ học sinh.
- Quản lý hồ sơ giáo viên.
- Quản lý điểm học tập của học sinh.
- Quản lý việc cấp phát văn bằng, chứng chỉ...
- Lập các báo cáo định kỳ vào các thời điểm đầu năm, giữa năm, cuối năm gửi về Sở Giáo dục và Đào tạo.

2. Đối với các phòng Giáo dục và Đào tạo cấp quận, huyện, thị.

Với việc phân cấp quản lý hiện nay thì vai trò của Phòng Giáo dục -Đào tạo ngày càng trở nên nặng nề, phạm vi quản lý của các Phòng Giáo dục - Đào tạo là quản lý trực tiếp với các cấp học: Mầm non, Tiểu học, PTCS và THCS. Với nhiệm vụ nặng nề đó, để đạt được hiệu quả trong công tác quản lý các Phòng Giáo dục và Đào tạo cần ứng dụng CNTT vào việc phục vụ quản lý Giáo dục và Đào tạo như sau:

- Quản lý đội ngũ cán bộ, giáo viên và công nhân viên ở tất cả các cấp học thuộc phạm vi mình quản lý.
- Quản lý hoạt động dạy, học tại các trường ở các cấp thuộc địa phương mình quản lý (nội dung, tiến độ, chất lượng hoạt động dạy và học ở tất cả các trường, các cấp học).
- Quản lý hệ thống cơ sở vật chất (trường, lớp, hệ thống thư viện, phòng tập thể dục, thể thao, bàn, ghế,…)
- Quản lý học sinh.
- Tìm kiếm và tra cứu học sinh ở trong phạm vi huyện, thị.
- Quản lý thi tốt nghiệp, tuyển sinh.
- Quản lý điểm.
- Quản lý và theo dõi việc đi học theo đúng độ tuổi.
- Công tác lập báo cáo đầu năm, giữa năm, cuối năm.

3.1.2.Yêu cầu tìm kiếm,khai thác thông tin quản lý từ các cơ quan chủ quản

1)Đối với Sở Giáo dục và Đào tạo:

- Đây là cơ quan quản lý cao nhất của ngành Giáo dục và Đào tạo trong phạm vi một tỉnh, thành phố, do đó các thông tin hai chiều có liên quan đến việc quản lý, chỉ đạo thực hiện việc dạy và học là đặc biệt quan trọng. Theo kỳ Sở Giáo dục và Đào tạo phải lập các báo cáo để gửi về Bộ Giáo dục và Đào tạo, các thông tin báo cáo chủ yếu trong báo cáo là các thông tin liên quan đến trường, lớp, học sinh, đội ngũ giáo viên, cơ sở vật chất… cụ thể là:

1. Trường :

- Số lượng các trường, số lượng từng loại hình đào tạo (trường chuyên, công lập, dân lập, bán công, tư thục)

2. Lớp :

- Số lượng lớp ở từng cấp học, bậc học.
- Số lượng các lớp học ngoại ngữ (Tiếng Anh, tiếng Nga, Tiếng Pháp, tiếng Trung)

3. Học sinh:

- Tổng số học sinh học ở các thời tại thời điểm báo cáo
- Số lượng học sinh nữ
- Số lượng học sinh người dân tộc
- Xếp loại học sinh về học lực và hạnh kiểm
- Tỉ lệ học sinh tốt nghiệp; Xếp loại tốt nghiệp : Giỏi, Khá và TB.
- Số học sinh tuyển mới
- Số học sinh lưu ban

4. Cán bộ, giáo viên:

Thông tin giáo viên, tổng số cán bộ, giáo viên, công nhân viên trong nhà trường, trong đó :

- Giáo viên trực tiếp giảng dạy (kể cả hợp đồng)
- Số GV người dân tộc.
- Trình độ đào tạo đạt trên chuẩn.
- Trình độ đào tạo đạt chuẩn.
- Lãnh đạo nhà trường: hiệu trưởng, các hiệu phó.
- Cán bộ phụ trách đoàn, đội.
- Nhân viên thư viện.
- Cán bộ phụ trách thí nghiệm.

5. Cơ sở vật chất :

- Số phòng học;
- Số thư viện;
- Số phòng tập thể dục thể thao.

6. Thông tin về chất lượng học sinh.

- Số lượng học sinh xếp loại theo từng khối, lớp.
- Xếp loại học lực theo các loại : Giỏi, Khá, TB, Yếu, Kém.
- Hạnh kiểm theo các mức: Tốt, Khá, TB, Yếu, Không xếp loại.

7. Thông tin có liên quan về thi tốt nghiệp.

- Thông tin về quản lý và cấp phát các loại bằng tốt nghiệp.
-Sự phân luồng của học sinh trong việc lựa chọn nghề nghiệp sau tốt nghiệp THCS và THPT.

8. Thông tin về tỉ lệ học sinh thi đỗ vào các trường đại học, cao đẳng.

9. Thông tin về số lượng học sinh đạt giải quốc gia, quốc tế.

10. Ngoài ra Sở Giáo dục và Đào tạo còn cần rất nhiều các thông tin phục vụ cho việc tra cứu và tìm kiếm dữ liệu trong phạm vi tỉnh mình.

2)Đối với Bộ Giáo dục và Đào tạo:

Bộ Giáo dục và Đào tạo là cơ quan cao nhất trực tiếp quản lý Giáo dục và Đào tạo, chịu trách nhiệm trước Đảng và nhân dân cả nước về chất lượng Giáo dục và Đào tạo. Hiện nay trước bối cảnh của xu thế toàn cầu hoá, Việt Nam cũng đang mở rộng quan hệ hợp tác với các nước và các tổ chức kinh tế quốc tế. Để có thể hoà nhập được với nền kinh tế - xã hội thế giới chúng ta cần có một nguồn nhân lực đủ trình độ có thể đáp ứng được đòi hỏi của xã hội. Trước tình hình đó yêu cầu của xã hội đặt ra đối với công tác quản lý, chỉ đạo của Bộ Giáo dục và Đào tạo với ngành càng trở nên cần thiết hơn bao giờ hết. Để hoàn thành được trách nhiệm của mình, Bộ Giáo dục và Đào tạo phải có những biện pháp nhất định trong việc tăng cường quản lý chỉ đạo chuyên môn, từng bước nâng cao chất lượng dạy và học ở các địa phương trong toàn quốc. Để thực hiện được các biện pháp điều hành Bộ Giáo dục và Đào tạo cần phải tăng cường trao đổi thông tin thường xuyên giữa Bộ Giáo dục và Đào tạo và các Sở Giáo dục và Đào tạo, đặc biệt là các thông tin ngược từ các Sở Giáo dục và Đào tạo về Bộ Giáo dục và Đào tạo là vô cùng quan trọng, các thông tin này giúp Bộ Giáo dục và Đào tạo có thể đưa ra được các giải pháp, biện pháp điều chỉnh cho phù hợp và kịp thời.

Quá trình chỉ đạo quản lý đối với ngành luôn có nhiều câu hỏi được đặt ra và yêu cầu phải được trả lời như :

- Việc đổi mới nội dung sách giáo khoa hiện nay của Bộ Giáo dục và Đào tạo đã đáp ứng được các yêu cầu đặt ra cũng như đáp ứng được các yêu cầu của xã hội giữa các vùng miền trên phạm vi cả nước hay chưa?

- Tỉ lệ học sinh tốt nghiệp các sở hàng năm trong cả nước là bao nhiêu.

- Biểu đồ xếp loại học sinh đỗ tốt nghiệp các cấp hàng năm như thế nào.

- Tỉ lệ đỗ tốt nghiệp của các học sinh người dân tộc thiểu số chiếm tỉ lệ bao nhiêu?

- Biểu đồ thể hiện các bậc điểm trong kỳ thi tốt nghiệp giữa các vùng trong phạm vi cả nước ?

- Tìm kiếm học sinh Nguyễn Hoà Bình trong cả nước.

- Tìm kiếm học sinh Nguyễn Văn An, sinh ngày 20/12/1975, tại Hoà Bình.

- Hàng năm các đơn vị, các trường cao đẳng và đại học mất rất nhiều thời gian và công sức để thực hiện công tác thanh, kiểm tra văn bằng chứng chỉ của tất cả các cán bộ hiện đang công tác trong khu vực biên chế nhà nước và của tất cả các học sinh, sinh viên chuẩn bị thi tốt nghiệp ra trường. Công tác này gặp nhiều khó khăn trong việc tìm kiếm hồ sơ ở các địa phương khác nhau, do hồ sơ thất lạc, ...

- Công tác quản lý đội ngũ giáo viên hiện nay cũng đang được quan tâm. Số lượng giáo viên đạt chuẩn và chưa đạt chuẩn hiện đang giảng dạy ở các cấp như thế nào. Số lượng cán bộ giáo viên đạt danh hiệu thi đua Giáo viên giỏi cấp tỉnh? Số lượng giáo viên người dân tộc thiểu số? Tỉ lệ số giáo viên là nữ hiện nay là bao nhiêu. Trong đội ngũ giáo viên hiện có bao nhiêu là Đảng viên ? …

- Công tác báo cáo thống kê về số lượng, chất lượng mạng tính định kỳ, đầu năm, cuối năm và giữa năm của các khối học, cấp học, bậc học.

- Số lượng các trường chuẩn quốc gia của các địa phương hiện nay là cơ sở để Chính phủ có kế hoạch đầu tư tài chính cho các tỉnh, thành trong cả nước trong việc xây dựng trường đạt chuẩn ?

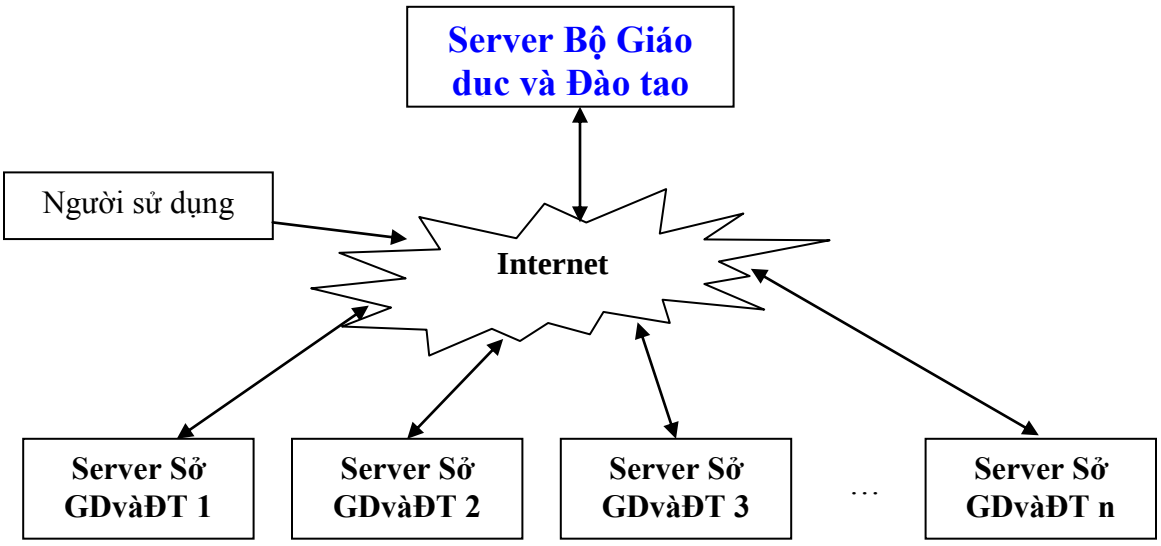
- Số lượng học sinh, giáo viên của các đơn vị là cơ sở để Chính phủ phân bổ ngân sách tài chính hàng năm cho ngành Giáo dục.

… Hàng loạt các câu hỏi khác tương tự như trên thường xuyên đặt ra trong quá trình chỉ đạo và quản lý của ngành Giáo dục và Đào tạo.

Để có được thông tin nhanh chóng về các vấn đề liên quan đến giáo dục cho các lãnh đạo, giúp cho các nhà lãnh đạo tăng cường công tác quản lý ngành Giáo dục và Đào tạo, đòi hỏi chúng ta phải xây dựng được một hệ thống thông tin phục vụ cho công tác quản lý Giáo dục và Đào tạo từ cấp Sở về cấp Bộ.

3.1.3.Mô hình hoá các yêu cầu

Từ các yêu cầu của các cấp Giáo dục trên chúng tôi đưa mô hình về tổ chức thông tin trong hệ thống thông tin giáo dục theo sơ đồ sau :



Mô hình khai thác và tìm kiếm thông tin trong cổng thông tin giáo dục

Ở trong sơ đồ này, máy chủ của Bộ Giáo dục và Đào tạo được nối với các máy chủ của các sở Giáo dục và Đào tạo qua môi trường Internet, CSDL của hệ thống được phân tán tại các server của các sở Giáo dục và Đào tạo. Nghĩa là, tại các server của các Sở Giáo dục và Đào tạo sẽ lưu toàn bộ dữ liệu quản lý về học sinh, trường, lớp và đội ngũ giáo viên, cán bộ công nhân viên trong phạm vi đơn vị mình quản lý. Khi cần tìm kiếm hay khai thác thông tin về một đối tượng nào đó của các Sở Giáo dục và Đào tạo,

tự động server của Bộ Giáo dục và Đào tạo sẽ tiến hành trao đổi thông tin với các server các Sở Giáo dục và Đào tạo.

Việc tìm kiếm thông tin trên Cổng thông tin giáo dục có thể được mô tả qua thí dụ sau: Một cơ quan cần tìm hiểu về thông tin của một học sinh, cơ quan này thông qua Internet kết nối đến cổng thông tin ngành Giáo dục và Đào tạo. Khi nhận được yêu cầu về tìm kiếm hoặc khai thác thông tin, lập tức máy chủ tại Bộ Giáo dục và Đào tạo sẽ yêu cầu các máy chủ của các Sở gửi về các thông tin cần thiết về, tập hợp kết quả và trả lại kết quả cho người sử dụng.

Việc lập báo cáo của các cơ quan quản lý giáo dục cũng tương tự như vậy, khi có yêu cầu về một loại báo cáo nào đó, người sử dụng chỉ cần lựa chọn các thông tin mà mình cần, máy chủ sẽ tự lấy các số liệu thống kê có liên quan từ các máy chủ của các Sở Giáo dục và Đào tạo. Rất nhanh chóng có ngay một báo cáo tổng hợp.

3.2.Tối ưu hoá hệ thống Cơ Sở Dữ Liệu

Từ các yêu cầu quản lý tại các cấp chúng ta thấy:

- 80% yêu cầu tìm kiếm và thống kê thông tin là được xử lý cục bộ tại máy chủ của đơn vị cơ sở (cấp Sở Giáo dục và Đào tạo).
- 15% yêu cầu tìm kiếm thống kê được xử lý phân tán trên hệ thống máy chủ của Bộ Giáo dục và Đào tạo.
- 5% là các yêu cầu khác.

Như vậy, hệ thống CSDL chi tiết sẽ chủ yếu sẽ được cài đặt tại các server Sở Giáo dục và Đào tạo, tại đây cũng cài đặt các CSDL thống kê nhằm đáp ứng nhu cầu thống kê tổng thể được nhanh.

Từ đó chúng tôi đã tiến hành tổ chức, phân tích, thiết kế xây dựng một hệ thống CSDL có liên quan phục vụ công tác quản lý Giáo dục, từ cơ quan Bộ Giáo dục và Đào tạo đến các Sở Giáo dục và Đào tạo với cấu trúc CSDL như sau :

3.2.1. Tại Bộ Giáo dục và Đào tạo :

Có CSDL **HOSOBO.MDF** với các bảng cấu trúc như sau:

Bảng 3.1. Thông tin về danh mục các Sở GD&ĐT (DMSOGD_DT).

STT	Tên trường	Mô tả
1	MA_SO	Mã Sở Giáo dục và Đào tạo
2	TEN_SO	Tên Sở Giáo dục và Đào tạo
3	URL	Địa chỉ máy chủ của các Sở Giáo dục và Đào tạo
4	DIEN_THOAI	Số điện thoại trực thi, báo cáo của các Sở
5	DIA_CHI	Địa chỉ liên hệ của các cơ sở Giáo dục và Đào tạo

Bảng 3.2 Thông tin về cấp học (caphoc)

STT	Tên trường	Mô tả
1	MA_CH	Mã cấp học
2	TEN_CH	Tên cấp học

Bảng 3.3. Thông tin về năm học (namhoc)

STT	Tên trường	Mô tả
1	MA_NH	Mã năm học
2	TEN_NAMHOC	Tên năm học

Bảng 3.4 thông tin về loại hình trường(truong)

STT	Tên trường	Mô tả
1	MA_DT	Mã đào tạo
2	TEN_DT	Tên loại hình đào tạo

Bảng 3.5 Thông tin về vùng ưu tiên(vung)

STT	Tên trường	Mô tả
1	MA_VUNG	Mã vùng ưu tiên
2	TEN_VUNG	Tên vùng ưu tiên

Bảng 3.6 Thông tin về dân tộc (dantoc)

STT	Tên trường	Mô tả
1	MA_DT	Mã dân tộc

2	TEN_DT	Tên dân tộc

Bảng 3.7 Thông tin về môn học (Mon hoc)

STT	Tên trường	Mô tả
1	MA_MON	Mã môn
2	TEN_MON	Tên môn học

Bảng 3.8 Thông tin về số liệu thống kê theo Sở Giáo dục và Đào tạo (THONGKE_THPT)

TT	Tên trường	Mô tả
1	MA_SO	Mã sở Giáo dục và Đào tạo
2	SL_TRUONG	Số lượng trường
3	TRUONG_CHUYEN	Số lượng trường chuyên
4	TRUONG_CONG LAP	Số lượng trường công lập
5	TRUONG_BAN CONG	Số lượng trường bán công
6	TRUONG_DANLAP	Số lượng trường dân lập
7	TRUONG_TUTHUC	Số lượng trường tư thục
8	SL_LOP	Số lượng lớp
9	LOP_CHUYEN	Số lượng lớp chuyên
10	LOP_CONG LAP	Số lượng lớp công lập
11	LOP_BAN CONG	Số lượng lớp bán công
12	LOP_DANLAP	Số lượng lớp dân lập
13	LOP_TUTHUC	Số lượng lớp tư thục
14	SL_HS	Số lượng học sinh
15	HS_CHUYEN	Số lượng học sinh trường chuyên
16	HS_CONG LAP	Số lượng học sinh trường công lập
17	HS_BAN CONG	Số lượng học sinh trường bán công
18	HS_DANLAP	Số lượng học sinh trường dân lập
19	HS_TUTHUC	Số lượng học sinh trường tư thục
20	SHS_NU	Số lượng học sinh nữ

21	SHS_DANTOC	Số học sinh người dân tộc ít người
22	SL_XA_01	Số xã thuộc vùng 01
23	SL_XA_02	Số xã thuộc vùng 02
24	SL_XA_03	Số xã thuộc vùng 03
25	SL_HK_YEU	Số lượng học sinh hạnh kiểm Yếu
26	SL_HK_TB	Số lượng học sinh hạnh kiểm Trung bình
27	SL_HK_KHA	Số lượng học sinh hạnh kiểm Khá
28	SL_HK_TOT	Số lượng học sinh hạnh kiểm Tốt
29	SL_HL_YEU	Số lượng học sinh học lực Yếu
30	SL_HL_TB	Số lượng học sinh học lực Trung bình
31	SL_HL_KHA	Số lượng học sinh học lực Khá
32	SL_HL_GIOI	Số lượng học sinh học lực Giỏi
33	SL_LOP01	Số lượng lớp năm thứ nhất (THCS lớp 5, THPT lớp 10)
34	SL_LOP02	Số lượng lớp năm thứ hai (6,11)
35	SL_LOP03	Số lượng lớp thứ ba (7,12)
36	SL_LOP04	Số lượng lớp thứ tư (8)
37	SL_LOP05	Số lượng lớp thứ năm (9)

Bảng 3.9.Thông tin về học sinh (hosohs)

Số TT	Tên trường	Mô tả
1	MAHS	Mã học sinh
2	HOCSINH_ID	Chỉ số học sinh.
3	HO_TEN	Họ tên học sinh
4	NGAY_SINH	Ngày sinh
5	NOI_SINH	Nơi sinh
6	HS_LOP	Học sinh lớp
7	DIA_CHI	Địa chỉ nhà riêng
8	MA_TRUONG	Mã trường
9	MA_DTOC	Mã dân tộc
10	GIOI_TINH	Giới tính

11	VUNG_MIEN	Vùng miền ưu tiên
12	DTB_TOAN	Điểm trung bình môn Toán
13	DTB_LY	Điểm trung bình môn Vật lý
14	DTB_HOA	Điểm trung bình môn Hoá học
15	DTB_SINH	Điểm trung bình môn Sinh học
16	DTB_VĂN	Điểm trung bình môn Văn
17	DTB_SỬ	Điểm trung bình môn Lịch sử
18	DTB_DIA	Điểm trung bình môn Địa lý
19	DTB_TIN	Điểm trung bình môn Tin học
20	DTB_TD	Điểm trung bình môn Thể dục
21	DTB_GDCD	Điểm trung bình môn Giáo dục công dân
22	DTB_NN	Điểm trung bình môn Tiếng nước ngoài
23	DTB_KH1	Điểm trung bình các môn học kỳ 1
24	DTB_HK2	Điểm trung bình các môn học kỳ 2
25	DTB_CN	Điểm trung bình môn chung cả năm
26	XL_HK1	Xếp loại hạnh kiểm học kỳ 1
27	XL_HL1	Xếp loại học lực học kỳ 1
28	XL_HK2	Xếp loại hạnh kiểm học kỳ 2
31	XL_HL2	Xếp loại học lực học kỳ 2
32	NTS	Năm Vào đầu cấp
33	NTN	Năm tốt nghiệp
34	XL_TN	Xếp loại tốt nghiệp
35	NAM_HOC	Năm học
36	KHEN_KY	Khen thưởng, kỷ luật
37	NHAN_XET	Nhận xét của GV chủ nhiệm về học sinh

Bảng 3.10 Thông tin về giáo viên (TK_Giaovien)

Số TT	Tên trường	Mô tả
1	MA_SO	Mã Sở Giáo dục và Đào tạo
2	TONG_SO	Tổng số cán bộ, cán bộ công nhân viên

3	TREN_CHUAN	Trình độ đạt trên chuẩn
4	DAT_CHUAN	Trình độ đạt chuẩn
5	CHUA_CHUAN	Trình độ chưa đạt chuẩn
6	SL_TOAN	Số lượng giáo viên Toán
7	SL_LY	Số lượng giáo viên Vật lý
8	SL_HOA	Số lượng giáo viên Hoá học
9	SL_SINH	Số lượng giáo viên Sinh học
10	SL_VAN	Số lượng giáo viên Văn
11	SL_SU	Số lượng giáo viên Lịch Sử
12	SL_DIA	Số lượng giáo viên Địa lý
13	SL_TIN	Số lượng giáo viên Tin
14	SL_GDCD	Số lượng giáo viên GDCD
15	SL_TD	Số lượng giáo viên Thể dục
16	SL_ANH	Số lượng giáo viên Tiếng Anh
17	SL_NGA	Số lượng giáo viên Tiếng Nga
18	SL_PHAP	Số lượng giáo viên Tiếng Pháp
19	SL_TRUNG	Số lượng giáo viên Tiếng Trung
20	HIEU_TRUONG	Số lượng hiệu trưởng
21	HIEU_PHO	Số lượng hiệu phó
22	DOAN_DOI	Số lượng cán bộ đoàn đội
23	THU_VIEN	Số lượng cán bộ thư viện
24	THI_NGHIEM	Số lượng cán bộ thí nghiệm
25	KT_NV	Số kỹ thuật viên kỹ thuật nghiệp vụ
26	PHUC_VỤ	Sè nh©n viª n phôc vô cßn l¹i

3.2.2. Tại Sở Giáo dục và Đào tạo :

Cã CSDL HOSOSO.MDF với các bảng cấu trúc được thiết kế như sau:

Bảng 3.11. Thông tin về danh mục các trường : (Truong)

STT	Tên trường	Mô tả
1	MA_TRUONG	Mã trường
2	TEN_TRUONG	Tên trường

3	MA_CH	Mã cấp học
4	MA_DT	Mã loại hình đào tạo của nhà trường
5	DIEN_THOAI	Số điện thoại thường trực thi hoặc lập báo cáo
6	DIA_CHI	Địa chỉ liên hệ.
7	TEN_HT	Tên hiệu trưởng
8	TEN_HP	Tên các hiệu phó

Bảng 3.12. Thông tin về cấp học (caphoc)

STT	Tên trường	Mô tả
1	MA_CH	Mã cấp học
2	TEN_CH	Tên cấp học

Bảng 3.13. Thông tin về năm học (namhoc)

STT	Tên trường	Mô tả
1	MA_NH	Mã cấp học
2	TEN_NAMGOC	Tªn n¨m häc

Bảng 3.14. Thông tin về loại hình trường (loaihinhtruong)

STT	Tên trường	Mô tả
1	MA_DT	Mã đào tạo
2	TEN_DT	Tên loại hình đào tạo

Bảng 3.15. Thông tin về vùng miền (vungut)

STT	Tªn tr-êng	Mô tả
1	MA_VUNG	Mã vùng ưu tiên
2	TEN_VUNG	Tên vùng - u tiên

Bảng 3.16. Thông tin về dân tộc (dantoc)

STT	Tên trường	Mô tả
1	MA_DT	Mã dân tộc

2	DAN_TOC	Tên dân tộc

Bảng 3.17.Thông tin về môn học (monhoc)

Số	Tên trường	Mô tả
1	MA_MON	Mã môn
2	TEN_MON	Tên môn

Bảng 3.18. Thông tin về thống kê theo Sở GD&ĐT (hososo)

Số TT	Tên trường	Mô tả
1	MA_SO	Mã Sở Giáo dục và Đào tạo .
2	SL_TRUONG	Số lượng trường
3	TRUONG_CHUYEN	Số lượng trường chuyên
4	TRUONG_CONG LAP	Số lượng trường công lập
5	TRUONG_BAN CONG	Số lượng trường bán công
6	TRUONG_DANLAP	Số lượng trường dân lập
7	TRUONG_TUTHUC	Số lượng trường tư thục
8	SL_LOP	Số lượng lớp
9	LOP_CHUYEN	Số lượng lớp chuyên
10	LOP_CONG LAP	Số lượng lớp công lập
11	LOP_BAN CONG	Số lượng lớp bán công
12	LOP_DANLAP	Số lượng lớp dân lập
13	LOP_TUTHUC	Số lượng lớp tư thục
14	SL_HS	Số lượng học sinh
15	HS_CHUYEN	Số lượng học sinh trường chuyên
16	HS_CONG LAP	Số lượng học sinh trường công lập
17	HS_BAN CONG	Số lượng học sinh trường bán công
18	HS_DANLAP	Số lượng học sinh trường dân lập
19	HS_TUTHUC	Số lượng học sinh trường tư thục
20	SHS_NU	Số lượng học sinh nữ
21	SHS_DANTOC	Số học sinh người dân tộc ít người

22	SL_XA_01	Số xã thuộc vùng 01
23	SL_XA_02	Số xã thuộc vùng 02
24	SL_XA_03	Số xã thuộc vùng 03
25	SL_HK_YEU	Số lượng học sinh hạnh kiểm Yếu
26	SL_HK_TB	Số lượng học sinh hạnh kiểm Trung bình
27	SL_HK_KHA	Số lượng học sinh hạnh kiểm Khá
28	SL_HK_TOT	Số lượng học sinh hạnh kiểm Tốt
29	SL_HL_YEU	Số lượng học sinh học lực Yếu
30	SL_HL_TB	Số lượng học sinh học lực Trung bình
31	SL_HL_KHA	Số lượng học sinh học lực Khá
32	SL_HL_GIOI	Số lượng học sinh học lực Giỏi
33	SL_LOP01	Số lượng lớp năm thứ nhất (THCS lớp 5, THPT lớp 10)
34	SL_LOP02	Số lượng lớp năm thứ hai (6,11)
35	SL_LOP03	Số lượng lớp thứ ba (7,12)
36	SL_LOP04	Số lượng lớp thứ tư (8)
37	SL_LOP05	Số lượng lớp thứ năm (9)

Bảng 3.19.Thông tin về học sinh (hosohs)

Số TT	Tên trường	Mô tả
1	MA_HS	Mã học sinh
2	HOCSINH_ID	Chỉ số học sinh.
3	HO_TEN	Họ tên học sinh
4	NGAY_SINH	Ngày sinh
5	NOI_SINH	Nơi sinh
6	HS_LOP	Học sinh lớp
7	DIA_CHI	Địa chỉ nhà riêng
8	MA_TRUONG	Mã trường
9	MA_DTOC	Mã dân tộc
10	GIOI_TINH	Giới tính

11	VUNG_MIEN	Vùng miền
12	DTB_TOAN	Điểm trung bình môn Toán
13	DTB_LY	Điểm trung bình môn Vật lý
14	DTB_HOA	Điểm trung bình môn Hoá học
15	DTB_SINH	Điểm trung bình môn Sinh học
16	DTB_VAN	Điểm trung bình môn Văn
17	DTB_SU	Điểm trung bình môn Lịch sử
18	DTB_DIA	Điểm trung bình môn Địa lý
19	DTB_TIN	Điểm trung bình môn Tin học
20	DTB_TD	Điểm trung bình môn Thể dục
21	DTB_GDCD	Điểm trung bình môn Giáo dục công dân
22	DTB_NN	Điểm trung bình môn Tiếng nước ngoài
23	DTB_KH1	Điểm trung bình các môn học kỳ 1
24	DTB_HK2	Điểm trung bình các môn học kỳ 2
25	DTB_CN	Điểm trung bình môn chung cả năm
26	XL_HK1	Xếp loại hạnh kiểm học kỳ 1
27	XL_HL1	Xếp loại học lực học kỳ 1
28	XL_HK2	Xếp loại hạnh kiểm học kỳ 2
29	XL_HL2	Xếp loại học lực học kỳ 2
30	NTS	Năm truyển sinh vào đầu cấp
31	NTN	Năm tốt nghiệp
32	XL_TN	Xếp loại tốt nghiệp
33	NAM_HOC	Năm học
34	BANG_TN	Số hiệu bằng tốt nghiệp được cấp
35	NHAN_XET	Nhận xét của giáo viên chủ nhiệm

Bảng 3.20. Thông tin về lượt truy cập của học sinh (luotID)

Số TT	Tên trường	Mô tả
1	MAHS	Mã học sinh
2	HOCSINH_ID	Chỉ số học sinh.

Bảng 3.21.Thông tin về giáo viên (HOSOGV)

Số TT	Tên trường	Mô tả
1	MA_GV	Mã giáo viên
2	GV_ID	Chỉ số giáo viên.
3	HO_TEN	Họ tên giáo viên
4	NGAY_SINH	Ngày sinh
5	NOI_SINH	Nơi sinh
6	DIA_CHI	Địa chỉ nhà riêng
7	MA_TRUONG_CT	Mã trường đang giảng dạy
8	MA_TRUONG_DT	Mã trường nơi đào tạo
9	Ma_TOC	Mã dân tộc
10	GIOI_TINH	Giới tính
11	NAM_CT	Năm bắt đầu vào biên chế chính thức
12	DH_TD	Danh hiệu thi đua
13	HS_LUONG	Hệ số lượng cơ bản hiện tại
14	MA_MON	Giảng dạy môn (mã môn)
15	CHUC_VU	Chức vụ đang đảm trách
16	VUNG_MIEN	Vùng miền
17	KT_KL	Khen thưởng, kỷ luật trong quá trình giảng dạy
18	TD_CM	Trình độ chuyên môn
19	Ghi_CHU	Ghi chú

3.3.Xây dựng chương trình

3.3.1.Các modul sẽ được xây dựng

Chương trình phục vụ tìm kiếm và xử lý thông tin giáo dục trong hệ thống thông tin giáo dục được chúng tôi chia làm các module sau :

1) Module cài đặt tại các Sở Giáo dục và Đào tạo :

Module này được cài đặt vào trang web của Sở Giáo dục và Đào tạo bao gồm các chức năng sau :

+ Cập nhật về danh sách các trường.

+ Cập nhật về danh sách học sinh từ các trường.

+ Cập nhật về danh sách giáo viên từ các trường.

+ Cập nhật các thông tin liên quan quản lý từ các trường.

2) Module cài đặt tại Bộ Giáo dục và Đào tạo, được thiết kế giao diện web bao gồm các chức năng sau :

+ Trang chủ: Đưa thông tin hoạt động của ngành Giáo dục và Đào tạo.

+ Tìm kiếm: Tìm kiếm thông tin về học sinh, giáo viên.

+ Thông tin quản lý :

- Cập nhật tự động số liệu thống kê từ các Sở Giáo dục và Đào tạo.
- Thông tin chi tiết về giáo viên của các Sở Giáo dục và Đào tạo.
- Thông tin chi tiết về học sinh, trường, lớp các Sở Giáo dục và Đào tạo.
- Lập báo cáo: lập các báo cáo của ngành Giáo dục và Đào tạo

+ Thảo luận :

- Thảo luận chuyên môn của giáo viên các bộ môn (định kỳ)
- Thảo luận quản lý của các lãnh đạo (họp trực tuyến).

+ Lịch làm việc : của Bộ và các Sở Giáo dục và Đào tạo

+ Thư viện điện tử: bài giảng, sách giáo khoa, thời khoá biểu, tranh ảnh đồ dùng dạy học và các sách tham khảo, E-learning ...

+ Văn bản chỉ đạo : Các văn bản hướng dẫn chỉ đạo ngành Giáo dục của Bộ Giáo dục và Đào tạo.

+ Hỗ trợ trực tuyến : Các hỗ trợ về công nghệ, giải đáp các thắc mắc trong tổ chức hoạt động của ngành...

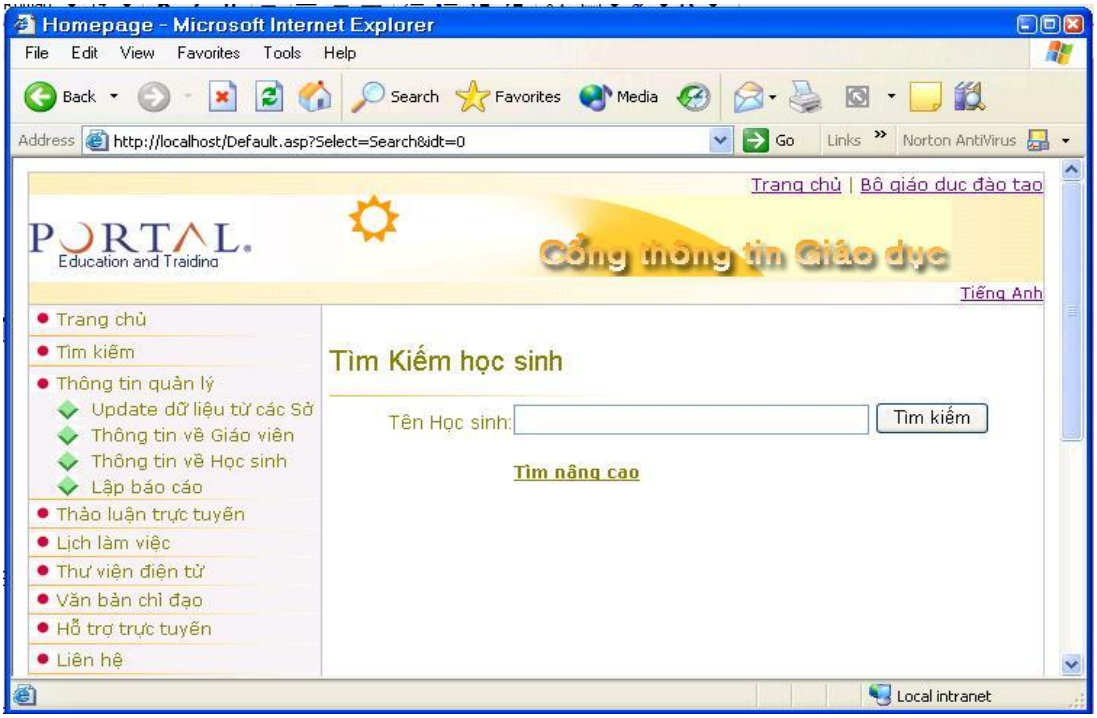
+ Liên hệ.

3.3.2. Giao diện cổng thông tin Giáo dục

a) Giao diện trang chủ của cổng thông tin giáo dục

Giao diện trang chủ Cổng thông tin giáo dục

b) Giao diện trang tìm kiếm học sinh theo tên

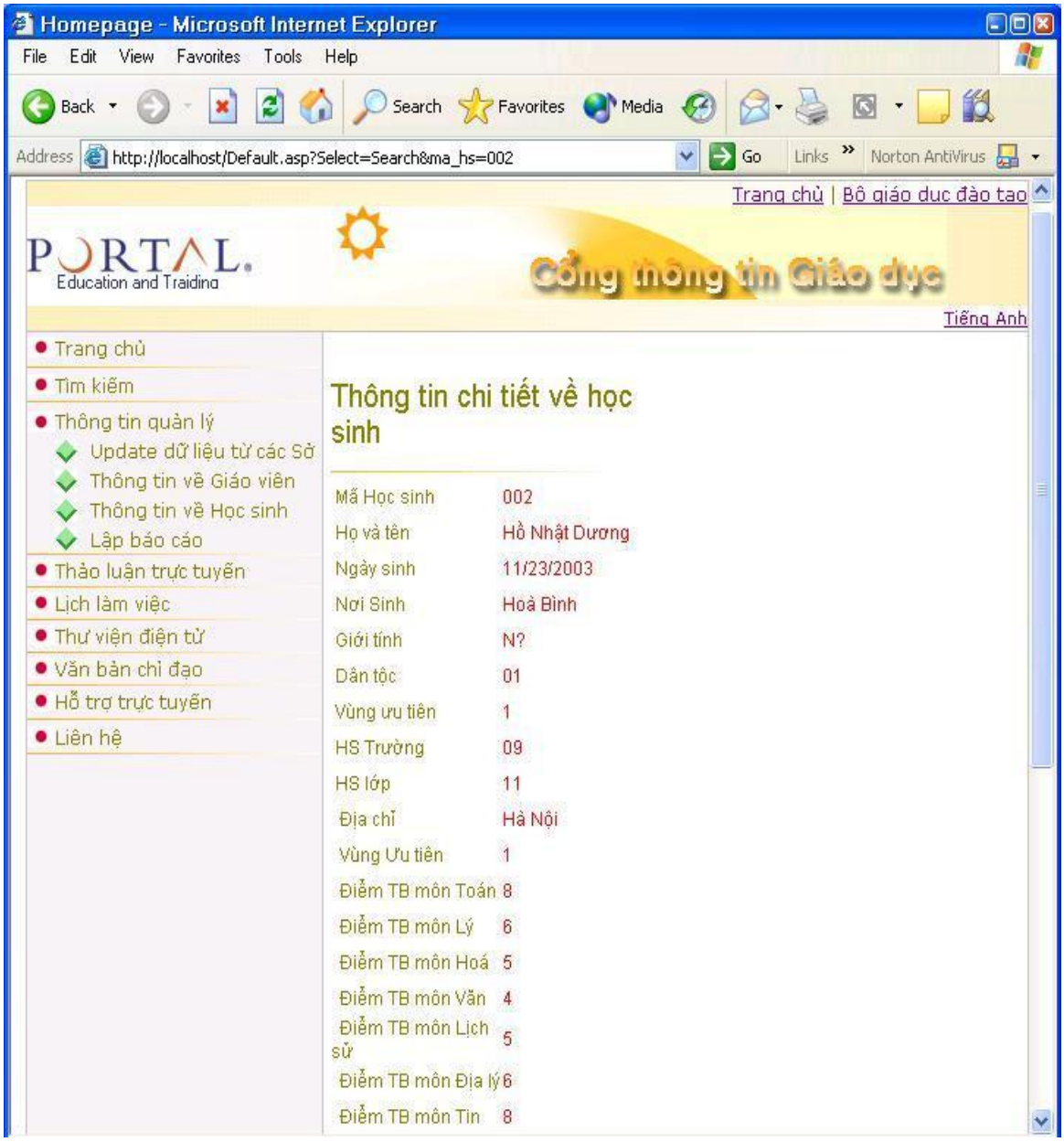


Giao diện trang tìm kiếm học sinh

c) Giao diện trang kết quả tìm kiếm học sinh theo tên

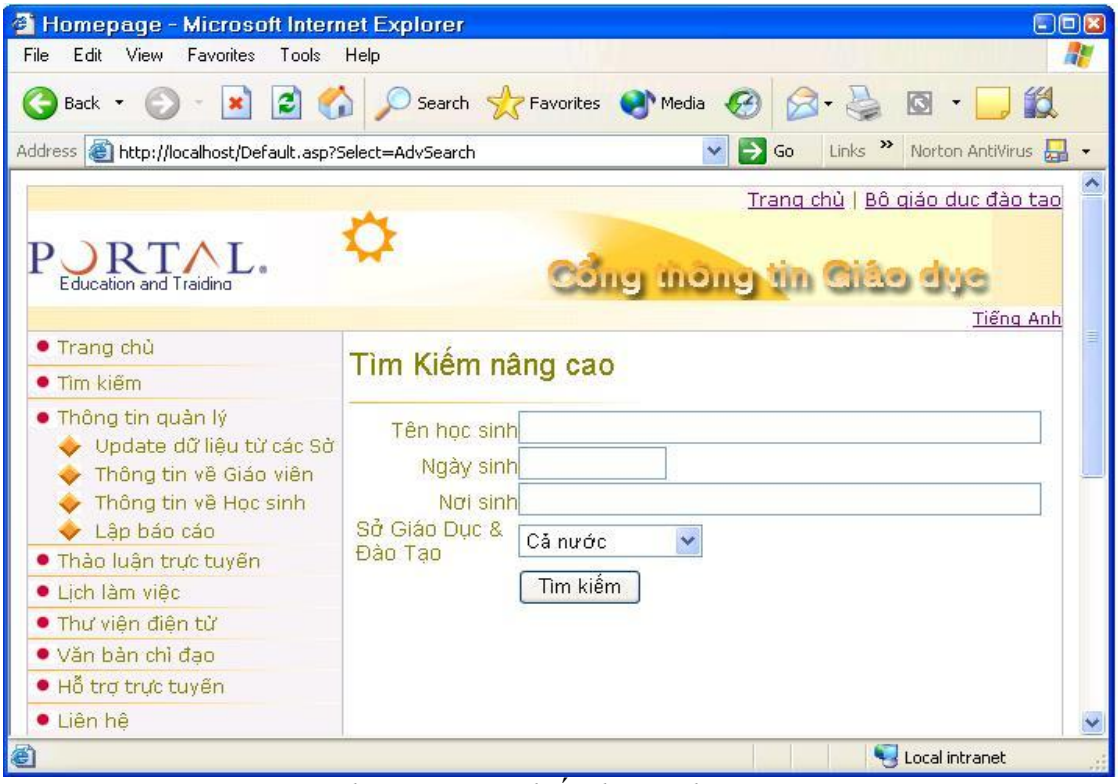
Giao diện trang tìm kiếm học sinh

d) Giao diện trang thông tin chi tiết về một học sinh



Giao diện trang thông tin chi tiết về một học sinh

e) Giao diện trang tìm kiếm nâng cao:

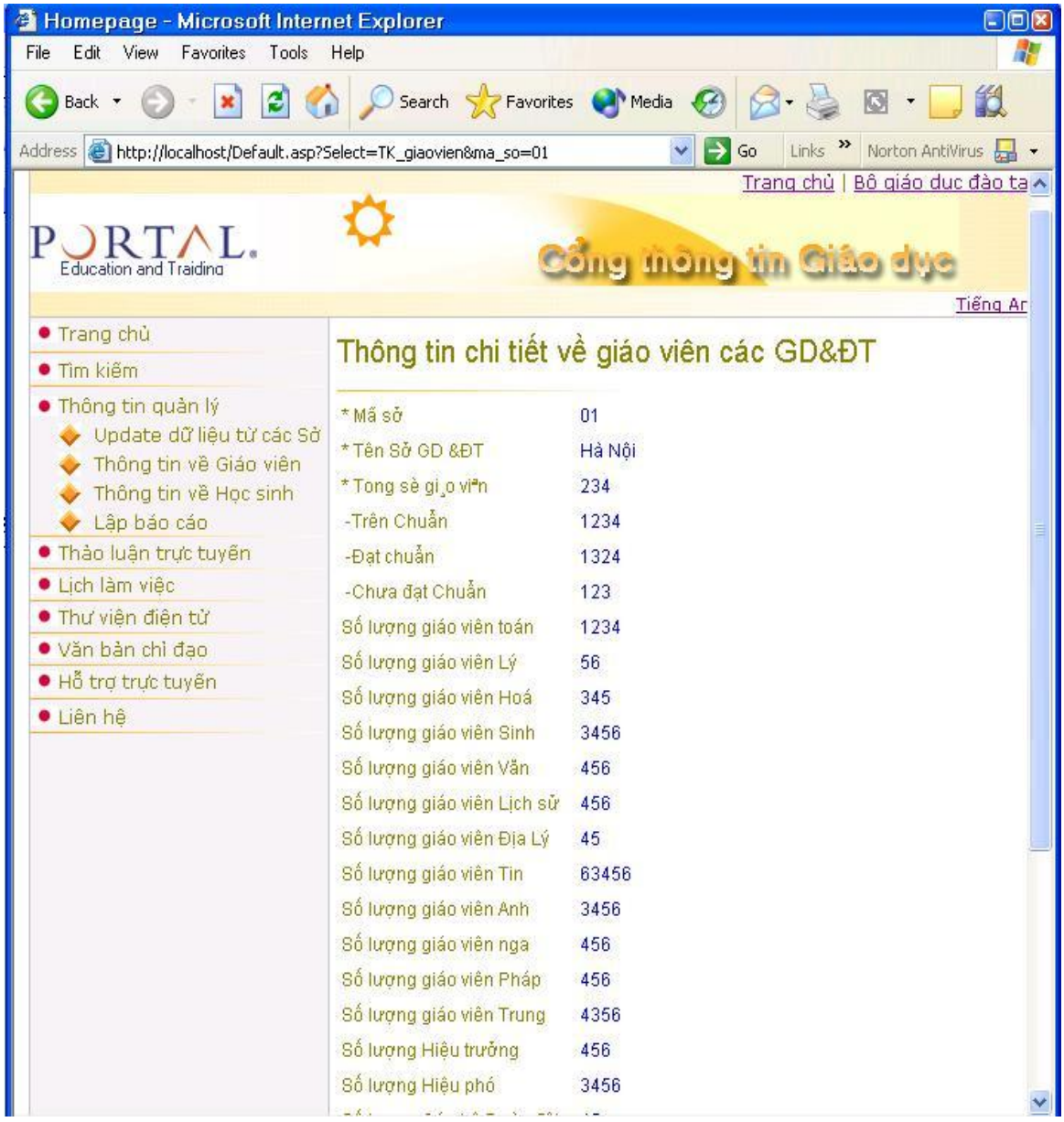


Giao diện trang tìm kiếm học sinh nâng cao

f) Giao diện trang thông tin về giáo viên của các Sở Giáo dục và Đào tạo trong phạm vi cả nước.

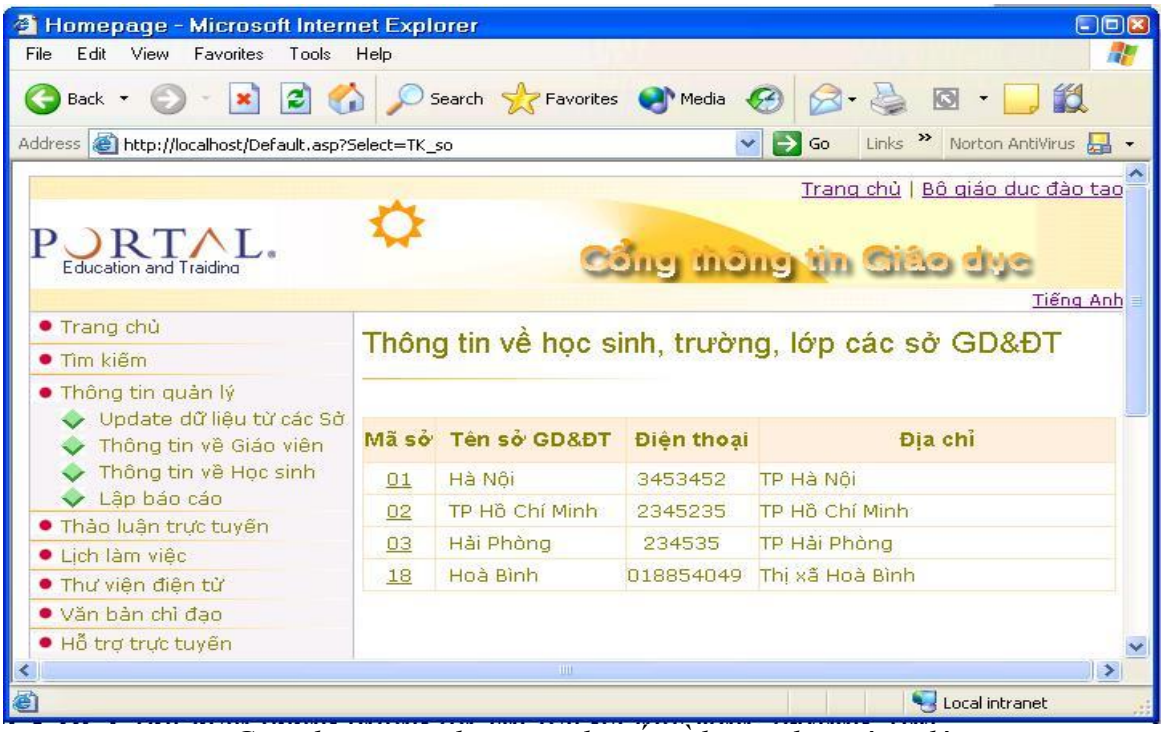
Giao diện trang khai thác thông tin giáo viên

g) Giao diện trang thông tin chi tiết về giáo viên của Sở Giáo dục và Đào tạo thành phố Hà Nội.



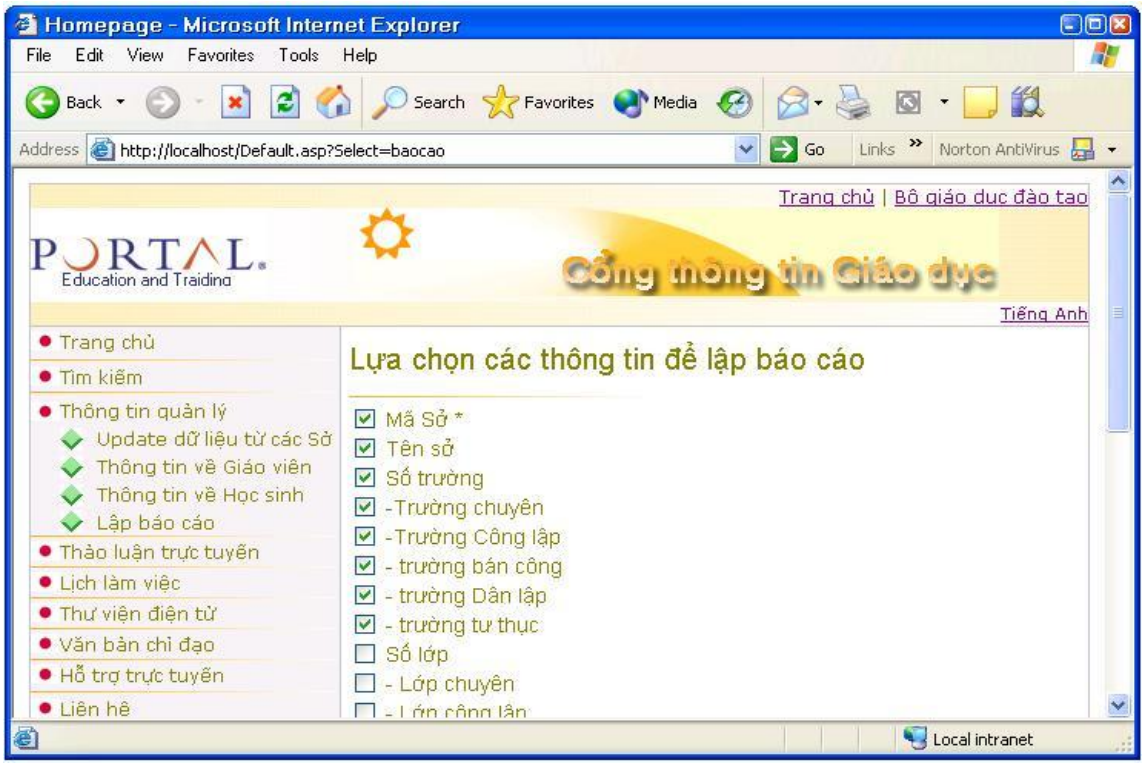
Giao diện trang thông tin tổng hợp về giáo viên
của sở Giáo dục và Đào tạo thành phố Hà Nội

h) Giao diện trang thông tin về trường, lớp của các Sở GD&ĐT



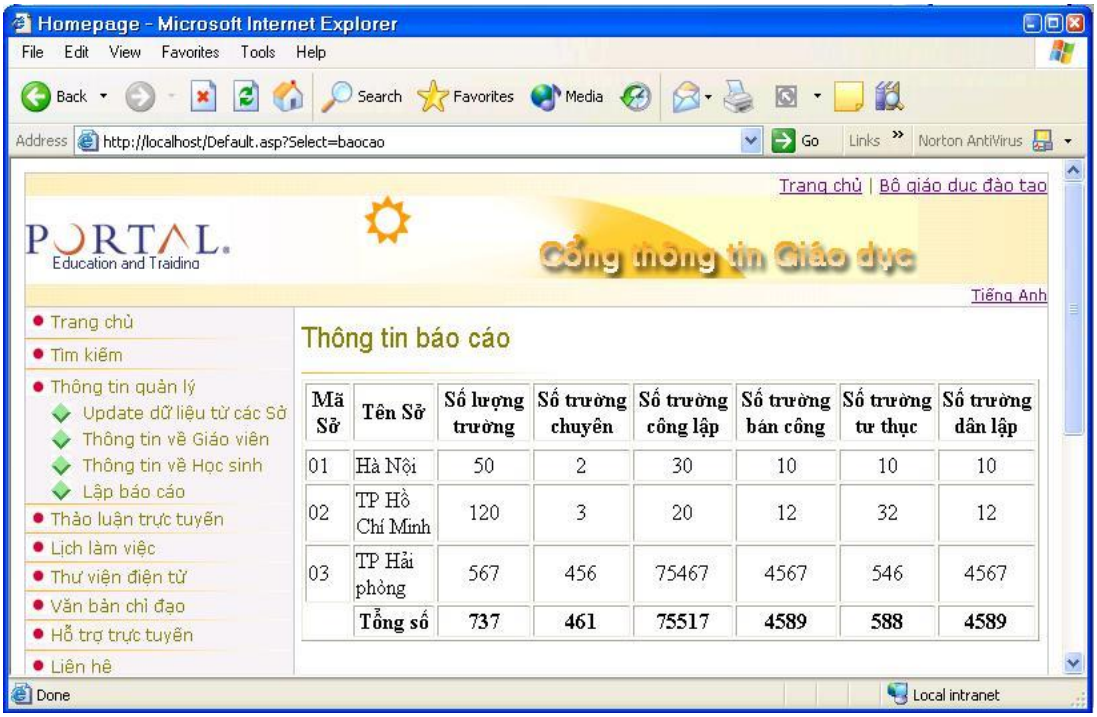
Giao diện trang thông tin chi tiết về học sinh, trường, lớp

j) Giao diện lựa chọn thông tin để lập báo cáo.



Giao diện trang lựa chọn thông tin để lập báo cáo

k) Giao diện báo cáo về các thông tin đã được chọn.



Mã Sở	Tên Sở	Số lượng trường	Số trường chuyên	Số trường công lập	Số trường bán công	Số trường tư thục	Số trường dân lập
01	Hà Nội	50	2	30	10	10	10
02	TP Hồ Chí Minh	120	3	20	12	32	12
03	TP Hải phòng	567	456	75467	4567	546	4567
	Tổng số	**737**	**461**	**75517**	**4589**	**588**	**4589**

Giao diện trang thông tin báo cáo

KẾT LUẬN

Với sự phát triển mạnh mẽ của CNTT, hiện nay điều kiện tiếp cận với thông tin thông qua mạng Internet đã trở nên dễ dàng với mọi người sử dụng. Việc xây dựng các hệ thống thông tin cho các ngành phục vụ nhu cầu quản lý điều hành hoạt động và khai thác dữ liệu đã trở nên cấp thiết và hiện thực hơn bao giờ hết. Kỹ thuật tiến tiến nhất để xây dựng hệ thống thông tin đó là Portal - Cổng thông tin điện tử. Luận văn đi sâu vào nghiên cứu một số vấn đề về khai thác và tìm kiếm dữ liệu thông qua công thông tin điện tử. Kết quả chủ yếu của luận văn bao gồm :

- Tổng quan về công nghệ Portal, đây là bước phát triển tiếp theo của Website. Công nghệ này cho phép người sử dụng khai thác dữ liệu và các dịch vụ cần thiết và "không cần phải đi đâu nữa".
- Nghiên cứu một số giải thuật tìm kiếm dữ liệu thực hiện trên cổng thông tin điện tử. Các kỹ thuật này liên quan đến tận dụng năng lực tính toán của hệ thống CSDL phân tán.
- Nghiên cứu thiết kế và tổ chức dữ liệu trên cổng thông tin điện tử ngành giáo dục. Việc tổ chức dữ liệu được chú trọng cho cấp cơ sở là cấp Sở Giáo dục và Đào tạo.

Các ứng dụng đã xây dựng đáp ứng được yêu cầu lớn trong việc tìm kiếm, thống kê thông tin, hỗ trợ việc lập báo cáo các hoạt động của ngành Giáo dục và Đào tạo. Là cơ sở để tăng cường công tác chỉ đạo các hoạt động Giáo dục từ Bộ Giáo dục và Đào tạo về các Sở Giáo dục và Đào tạo được nhanh chóng, kịp thời; góp phần từng bước nâng cao chất lượng Giáo dục và Đào tạo tại Việt Nam; thúc đẩy tốc độ hội nhập của nước ta với nền kinh tế thế giới và đẩy nhanh sự nghiệp công nghiệp hoá, hiện đại hoá đất nước. Góp phần vào công cuộc xây dựng Chính phủ điện tử nước nhà.

Để cổng thông tin giáo dục thực sự là nơi cung cấp thông tin của ngành Giáo dục và Đào tạo, góp phần vào việc cải cách thủ tục hành chính; là công cụ để đổi mới, tăng cường công tác quản lý Giáo dục và Đào tạo, đề tài cần được tiếp tục phát triển theo các hướng như sau :

- Tiếp tục hoàn thiện các module đã được xây dựng.
- Tiếp tục phát triển thêm các dịch vụ mới của cổng thông tin, làm phong phú nội dung thông tin trên cổng thông tin giáo dục.
- Đề nghị với Bộ Giáo dục và Đào tạo cho phép triển khai xây dựng Cổng thông tin giáo dục.
- Tiến hành kết nối Cổng thông tin giáo dục tại Bộ Giáo dục và Đào tạo với các Sở Giáo dục và Đào tạo trong cả nước.
- Kết nối Cổng thông tin giáo dục đến máy chủ của các trường đại học, cao đẳng trong phạm vi cả nước.

TÀI LIỆU THAM KHẢO

Tiếng việt

1. Ban điều hành đề án 112 (2004), *Giáo trình thiết kế và quản trị web, tổng quan Portal*.
2. Bộ Giáo dục và Đào tạo (2004), *Sổ tay h□ớng dẫn nghiệp vụ thống kê Giáo dục và Đào tạo*.
3. Lê Hữu Đạt, Nguyễn Ph- ơng Lan (2001), *ASP 3.0 và ASP.NET*, NXB Lao động – Xã hội.
4. HTTP://www.manguon.com/Ebooks/Details/EBO 034226135812, *Ngôn ngữ ASP*
5. Jeffrey D. Ullman (1998), *Nguyên lý các hệ CSDL và cơ sở tri thức,* NXB Thống kê.
6. Nguyễn ph- ơng Lan (2003), *XML nền tảng và ứng dụng*, NXB Lao động – Xã hội.
7. Tr- ơng Công Lục, Mai Hoàng Quý (2000), *Thiết kế và xuất bản trang web với HTML,* NXB Thống kê.
8. Nhóm tác giả ĐHBK Hà Nội (2002), *Th□ơng mại điện tử với VB, SQL 2000, MTS ASP Database,* NXB Trẻ
9. Đỗ Trung Tuấn (2004), *Cơ sở dữ liệu*, NXB Giáo dục
10. Nguyễn Bá T- ờng (2001), *Cơ sở dữ liệu lý thuyết và thực hành,* NXB Khoa học kỹ thuật Hà Nội

Tiếng Anh

11. Al Mc Kinnon and Mc Kinnon (2003), *XML*
12. Arvind Arasu, Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke, Sriram Raghavan (2000). *Searching the web*. Technical Report, Computer Science Department, Stanford University.
13. HTTP://www.Redbooks.ibm.com (2003), *Architecting Portal Solution*.
14. HTTP://www.Microsoft.com/uk/windowsserversystem/portals/what-is/default.mspx, *What is a Portal*
15. HTTP://www.vninformatics.com/portal/news/database, *SQL Server*
16. HTTP://www.nottingham.ac.uk/portals2002, *Conference the Portal*
17. HTTP://www.fair-portal.hull.ac.uk/, *The Portal project*
18. HTTP://www.xml.com, *XML*
19. HTTP://www.Doublincore.com
20. HTTP://www.Redbooks.ibm.com (2004), *XML for DB2 Information Integration*.

Mục lục

trang