

SINH VIÊN KHOA CNTT - ĐHDL VĂN LANG VỚI NHỮNG HƯỚNG KHOA HỌC - CÔNG NGHỆ MỚI

TS. Nguyễn Lâm

Hoạt động nghiên cứu khoa học sinh viên khoa CNTT-ĐHDL VL được thực hiện theo hai phương thức chủ yếu: làm đề tài tốt nghiệp và làm đề tài nghiên cứu được Nhà trường phê duyệt theo đăng ký. Đề tài tốt nghiệp thường tập trung vào năm cuối, còn đề tài nghiên cứu theo kế hoạch Trường thì có thể thực hiện ở những năm trước.

Nhìn chung các đề tài xét theo khía cạnh nào đó đều có những vấn đề liên quan đến các hướng khoa học công nghệ mới. Sau đây xin trình bày một số hướng.

1. Hệ thống thông tin quản lý. (quản lý doanh nghiệp). Số đề tài theo hướng này chiếm khoảng 36%, liên quan đến các hệ thống : MRP (Material Resource Planning), MRP II (Manufacturing Resource Planning), ERP (Enterprise Resource Planning). Đối với ERP thì các đề tài chỉ mới tiếp cận đến các hệ có chức năng riêng rẽ như: quản lý nhân lực, quản lý tiền lương, quản lý kế toán, tài chính... Cũng có đề tài liên kết một số chức năng, nhưng chưa có đề tài nào liên kết tất cả chức năng của kế hoạch nguồn lực doanh nghiệp.

Hệ thống tin ERP đã quá phổ biến. Các doanh nghiệp lớn hiện nay đều có hệ ERP của mình. Để xây dựng và cài đặt cho hệ ERP hoạt động có hiệu quả các doanh nghiệp phải tốn hàng triệu USD. (Có doanh nghiệp chi phí cho chuyên gia tư vấn trong thời gian cài đặt thử nghiệm mỗi ngày từ 800 đến 1000 USD và phải mất hàng nghìn ngày công). Tuy nhiên sinh viên khoa CNTT vẫn có thể tiếp cận đến các hệ ERP cho các doanh nghiệp vừa và nhỏ, có thể giải quyết một khâu nào đó nằm trong hệ thống hoàn chỉnh. Cũng không loại trừ khả năng xây dựng hệ ERM (Enterprise Resource Management). Đối với những hệ này yêu cầu không dừng lại ở lập kế hoạch mà phải ra quyết định. Những kiến thức liên quan "Ra quyết định" cần được trang bị như Lý thuyết ra quyết định, Data Mining... Từ đây trong chương trình đào tạo không chỉ dừng lại ở phân tích, thiết kế, cài đặt mà còn phải trang bị thêm tri thức để xử lý, ra quyết định.

2. Microsoft . NET. Khoảng 15% đề tài có sử dụng công nghệ này. Với DOS, rồi Windows và bây giờ đang là .NET, Microsoft tập trung vào xây dựng các công cụ hỗ trợ phát triển ứng dụng. NET cung cấp một khung cho ứng dụng (Application Framework), là một nền tảng trên đó có thể xây dựng các giải pháp và các dịch vụ WEB, một nền tảng giải phóng các ràng buộc (và ngay chính nó cũng được giải phóng khỏi Microsoft Windows về mặt kỹ thuật). Nói cách khác. NET là một bộ công cụ để xây dựng các ứng dụng và các dịch vụ hoạt động không phụ thuộc vào một nền tảng (Platform) nào. Microsoft xây dựng phần mềm Visual Studio.NET. Nó có 3 phiên bản:

- Visual Studio . NET Enterprise Architect dành cho các kiến trúc sư hệ thống phát triển cơ sở hạ tầng.
- Visual Studio .NET Enterprise Developer dành cho phát triển xây dựng các dịch vụ WEB XML nhằm chủ yếu vào các ứng dụng doanh nghiệp.
- Visual Studio .NET Professional dành cho cá nhân các nhà phát triển xây dựng các dịch vụ WEB XML.

Sinh viên Khoa CNTT –ĐHDLVL ứng dụng các công cụ này đã nhiều năm nay, đầu tiên là ASP.NET. Gần đây, đặc biệt là làm tốt nghiệp cho khoá 5 có nhiều đề tài liên quan đến Microsoft . NET.

Microsoft .NET cung cấp một nền tảng chuẩn cho phép tái sử dụng, tận dụng những ưu điểm của những ngôn ngữ đã có. Điều này là một lợi thế so với J2EE (Java Enterprise Edition) của Sun Microsystems. J2EE tuy hướng đến các ứng dụng WEB nhưng lại bỏ qua những kiến thức đã thu nạp trước để bắt đầu một nền tảng mới. Visual Studio . NET và .NET Framework SDK (Software Development Kit) hỗ trợ 8 ngôn ngữ: Anh, Pháp, Đức, Ý, Tây ban Nha, Nhật, Trung quốc, Triều tiên và hỗ trợ đến 26 ngôn ngữ lập trình khác nhau, trong đó có: C++, C#, JAVA, VB, PERL, OZ, ADA, OBREON, EIFFEL, MERCURY, FORTRAN, DELPHI...

3. Search Engine (S.E.). Số đề tài làm theo hướng này còn ít, chiếm khoảng 1%, nhưng đây là một hướng rất cần khuyến khích sinh viên chọn lựa., đặc biệt là nghiên cứu Search Engine hỗ trợ tiếng Việt.

Để tìm kiếm thông tin trên internet cần phải có những công cụ tìm kiếm. Đó là các Search Engine. Người khai thác thông tin trên internet khá quen thuộc với các Search Engine như: GOOGLE , ALTA VISTA , LYCOS , NETNAM , VINASEEK (có hỗ trợ tiếng Việt) . Sau đây trình bày một số nét cơ bản của một số Search Engine.

- **GOOGLE.** Là một trong những Search Engine được người dùng ưa chuộng. GOOGLE sử dụng cấu trúc các siêu liên kết để tính độ phổ biến cho mỗi trang Web và khai thác các siêu liên kết để cải thiện kết quả tìm kiếm. Các trang Web tìm được sẽ được lưu trữ trong bộ nhớ của Server dưới dạng nén. Việc lập chỉ mục thực hiện bởi bộ lập chỉ mục và bộ sắp xếp. Bộ lập chỉ mục thực hiện các chức năng như đọc dữ liệu , giải nén và phân tích tài liệu. Từ tập tài liệu đã phân tích rút ra tần số xuất hiện các Từ. Kết quả được sắp xếp tự động. Thuật toán GOOGLE sử dụng là thuật toán tính độ phổ biến PAGE RANK . Đây là một phần mềm phân loại các trang Web được phát triển bởi Larry Page và Sergei Brin ở đại học Stanford. Page Rank sử dụng cấu trúc liên kết các trang Web như một giá trị khởi động của trang đó, từ đó xác định trọng số (độ quan trọng) của trang . Website quan trọng sẽ có trọng số cao nhưng không phải vì thế mà nó luôn luôn có giá trị. Nếu nội dung không phù hợp với yêu cầu truy vấn sẽ không có giá trị. Vì thế Google đã kết hợp thuật toán Page Rank với kỹ thuật so khớp từ khoá để tìm ra các trang vừa quan trọng lại vừa phù hợp với nội dung tìm kiếm. Kết quả phù hợp nhất cho truy vấn không chỉ dựa vào số lần từ tìm kiếm xuất hiện mà còn xem xét đến nội dung trang và nội dung các trang liên kết đến nó. GOOGLE hỗ trợ YAHOO chức năng tìm kiếm.

- **ALTA VISTA .** Search Engine này thực hiện được với 25 ngôn ngữ. Là một Search Engine mạnh về tìm kiếm theo từ khoá. Alta Vista cũng tổ chức dữ liệu theo thư mục như các S.E. khác nhưng cùng với kết quả tìm được nó còn xem xét thêm thông qua một số câu hỏi liên quan đến vấn đề tìm kiếm. Alta Vista xây dựng một công cụ là công cụ Babelfish cho phép dịch từng câu hay cả trang Web từ ngôn ngữ này sang ngôn ngữ khác. Phần mềm Scooter của Alta Vista thì đảm nhận việc kiểm tra , duy trì các chỉ mục. Nếu siêu liên kết đến trang Web nào không hoạt động thì trang đó sẽ bị loại trừ khỏi danh sách chỉ mục. Scooter phát đi cùng một lúc hàng ngàn tiến trình , chuyển thông tin cho hệ thống lập chỉ mục. Một ngày Scooter duyệt hàng chục triệu trang Web. Alta Vista không những chỉ quan tâm dữ liệu Metatag (câu lệnh nhúng vào Header của trang Web) mà còn quan tâm đến mọi từ trong trang Web. Alta Vista thực hiện lập chỉ mục trên toàn bộ văn bản (Full-Text Indexing) , dữ liệu không phân hạng, không bảo trì, các tập tin đều không có cấu trúc và cũng không có thứ tự. Đây là những tính năng độc đáo của Alta Vista .

- **NETNAM.** Là một S.E. có hỗ trợ tiếng Việt. Thực ra Netnam nhằm mục tiêu chính là tiếng Việt. Hệ thống được thiết kế theo kiến trúc xử lý song song với 3 khối chức năng chính: thu thập thông tin, chuyển đổi dữ liệu và tổ chức cơ sở dữ liệu.

- **VINASEEK .** Cũng là một S.E. hỗ trợ tiếng Việt, cho phép tìm kiếm và hiển thị theo bất kỳ bảng mã nào. S.E. là một hướng nghiên cứu rộng mở . Hiện đang có nhiều đề tài các cấp ở nước ta nghiên cứu xây dựng các S.E. hỗ trợ tiếng Việt. Sinh viên Khoa CNTT-ĐHDLVL có thể tham gia vào một số khâu nào đó trong quá trình này. Chẳng hạn tìm kiếm theo ngữ nghĩa (tiếng Việt , có thể tập trung vào một lĩnh vực nào đó) . Biểu diễn ngữ nghĩa có thể xem như một bài toán con của biểu diễn tri thức . Ở đây có những vấn đề như các dạng quan hệ ngữ nghĩa , cách chúng được tổ chức thành hệ thống trong các hệ biểu diễn ngữ nghĩa , từ điển ngữ nghĩa (WORDNET) , lý thuyết các hệ thống biểu diễn ngữ nghĩa (ONTOLOGY) . Những vấn đề này tuy trừu tượng nhưng sinh viên Khoa CNTT-ĐHDLVL không thể vì trừu tượng mà lẩn tránh khi nhu cầu khai thác thông tin trên internet ngày càng phát triển.

4. Phần mềm nguồn mở. PMNM.

Đây là một hướng nghiên cứu đang được khuyến khích ở nước ta. Trên thế giới thì PMNM trở thành một hiện tượng đặc biệt. PMNM là những phần mềm được cung cấp dưới dạng mã nhị phân (Binary Code) và cả mã nguồn (Source Code) , thường là miễn phí về bản quyền, người dùng có quyền sửa đổi, cải tiến, phát triển , nâng cấp chỉ cần tuân theo một số nguyên tắc chung. Phần mềm nguồn mở là một trào lưu tin cậy đang phát triển rất nhanh: 90% máy chủ tên miền ở Hoa kỳ , 70% máy chủ thư tín điện tử, 60% máy chủ Web trên thế giới đã dùng phần mềm nguồn mở. Đến giữa năm 2001 đã có 50% số doanh nghiệp toàn cầu sử dụng phần mềm nguồn mở . Tháng 6/2001 Bộ Văn hoá Pháp đã chuyển Hệ điều hành 400 máy chủ sang Linux , Tổng cục Thuế quốc gia đã chuyển Hệ điều hành 950 máy chủ sang hệ điều hành nguồn mở. Trung quốc, Mêhicô, Hàn quốc đã chuẩn y các biện pháp để chuyển hệ thống máy tính của tất cả các cơ quan hành chính sang Linux . Chính phủ Đức có lệnh cấm dùng sản phẩm Microsoft trong các hệ thống máy tính " nhạy cảm ". Cộng đồng châu Âu khuyến cáo các tổ chức của cộng đồng và các cơ quan công quyền

của các quốc gia thành viên "**khuyến khích các dự án dùng phần mềm có mã nguồn mở, vì đó là cách duy nhất bảo đảm trong phần mềm không có các " cửa sau"(back doors)**".

Sinh viên Khoa CNTT-ĐHDLVL được học sử dụng hệ điều hành Linux ngay từ khoá 1 (1995-2000) và ngày càng mở rộng. Nhiều đề tài nghiên cứu liên quan đến hệ điều hành Linux được thực hiện , trong đó đề tài "chương trình bộ gõ tiếng Việt trên Linux " đã được nghiệm thu.

Những nghiên cứu phát triển các ứng dụng trên hệ điều hành Linux được Khoa khuyến khích .

Những sản phẩm PMNM sau đây đã khá phổ biến:

Về hệ điều hành thì ngoài Linux còn có NETBSD và FreeBSD. Hai hệ sau do Đại học Berkeley-Hoa kỳ phát triển.

Về ngôn ngữ lập trình thì có Perl, Python...

Các máy chủ Web: Apache , CERN...

Các máy chủ Proxy: Squid, CERN...

Các hệ quản trị CSDL quan hệ : MySQL. PostgreSQL...

Các hệ phát triển trang Web : HTML, XML...

Phần mềm SAMBA cho phép mô phỏng Windows 2000 và Windows NT cho phần quản lý tệp.

Phần mềm LDAP quản trị thông tin trên mạng.

5. Phần mềm tích hợp (cho các trung tâm tích hợp dữ liệu, chẳng hạn cho chính phủ điện tử, thương mại điện tử) – Enterprise Information Integration-EII

Việc tích hợp dữ liệu không ngoài mục đích hỗ trợ cho việc phân tích dữ liệu trực tuyến (OLAP, ROLAP) , tạo công cụ hỗ trợ quyết định, đảm bảo dữ liệu cho các ứng dụng, cho các cổng điện tử và dịch vụ Web. Những nguồn dữ liệu rất đa dạng , nhu cầu khai thác đòi hỏi độ tăng cấp (Scalability), tin cậy (Reliability), hiệu xuất (Performance), toàn vẹn dữ liệu (Data Integrity) cao, chức năng phong phú về bảo mật , quản trị , liên kết với cổng điện tử, hỗ trợ các chuẩn lập trình như: .NET, J2EE, EJB, Webservice, Servlet, Portal, Application Server.... CSDL tích hợp sẽ được sử dụng phát triển ứng dụng liên ngành như cổng điện tử và các ứng dụng dịch vụ công (trong chính phủ điện tử).

EII có sự khác biệt cơ bản với Data Warehouse (truyền thống). Data Warehouse tập trung dữ liệu theo định kỳ vào CSDL, hỗ trợ cho việc phân tích. Còn EII cũng tích hợp dữ liệu nhưng với các chức năng phong phú hơn: có dữ liệu không truy xuất từ lưu trữ ở CSDL của EII mà được truy xuất trực tiếp từ CSDL nguồn khi cần .Như vậy EII có khả năng cung cấp dữ liệu trực tuyến . EII còn có khả năng phân phối dữ liệu đến các CSDL nguồn (Replication), hỗ trợ công cụ quản lý danh bạ của Metadata , tạo các công cụ , tiện ích để xây dựng và vận hành các ứng dụng.

Sau đây điếm qua một số sản phẩm EII.

-MICROSOFT: Microsoft Host Integration Server 2000 cho phép CSDL tích hợp làm việc với các CSDL khác như DB2, Oracle, Informix, SQL Server, AS/400 và VSAM qua ODBC, OLEDB. Tuy nhiên nó cũng chỉ là một CSDL quan hệ có thể kết nối với các CSDL khác. Cho đến nay Microsoft vẫn chưa có sản phẩm cho giải pháp EII.

-ORACLE. Oracle 9i Integration có thể tích hợp dữ liệu từ các CSDL dạng khác như DB2, Informix , SQL Server, Sybase. Sản phẩm này vẫn chưa thực sự là một EII.

-OPEN SOURCE. Chưa phải là một sản phẩm EII.

-IBM. Đến nay IBM cung cấp giải pháp EII mở và hoàn thiện nhất. Giải pháp của IBM bao gồm DB2 Information Integrator-DB2II (cho dữ liệu có cấu trúc), DB2 Information Integrator for Content (cho dữ liệu phi cấu trúc).

DB2II dựa trên các chuẩn mở và có khả năng kết nối với các loại Webserver, Portal, Application Server... khác nhau. Nó vừa là CSDL tích hợp vừa là CSDL phân tán ,có khả năng phân phối dữ liệu trở lại cho các CSDL nguồn.

6. Cổng thông tin. Vấn đề tích hợp thông tin và ứng dụng Web đang đặt ra cấp bách khi ở nước ta đã có nhiều Website cho các ngành , các tổ chức, doanh nghiệp và khi nhu cầu tích hợp thông tin phục vụ cho nhiều hoạt động khác nhau trong xã hội ngày càng cao.Cổng thông tin là một công nghệ đáp ứng yêu cầu đó. Phần mềm cổng thông tin mã nguồn mở uPortal là một trong các giải pháp tích hợp thông tin và ứng dụng mạng trong một Website , với các cơ chế quản lý và xác thực người dùng hiệu quả, thích hợp cho mọi tổ chức với quy mô từ nhỏ đến lớn. Phần mềm uPortal là sản phẩm phần mềm nguồn mở , được cung cấp miễn phí với đầy đủ mã nguồn , nhưng nhiều ứng dụng (Channel) trên nền uPortal lại là các sản phẩm thương mại.

Tư tưởng chủ đạo của uPortal là tạo ra một giao tiếp (Interface) chung cho thông tin và ứng dụng. Để thực hiện điều này tất cả thông tin xuất ra cho các channel đều phải được cho dưới dạng XML . Người dùng thông qua trình duyệt Web của mình chọn lựa các kiểu trình bày thông tin khác nhau sử dụng các thẻ định dạng XSL (Stylesheets) để có được các trang thông tin trình bày theo yêu cầu.

Một số phần mềm ở nước ta trên nền công nghệ này đã xây dựng các sản phẩm riêng như iPortal (Unisoft- Compotech, sử dụng ở trường đại học Quốc gia TP.HCM) , VPortal (Vietsoftware, sử dụng cho cổng thông tin tp. Hà Nội) , VnPortal (Đan phong, sử dụng trong phần mềm VnOffice).

Theo hướng công nghệ này rất nhiều vấn đề đặt ra để nghiên cứu ứng dụng . Sinh viên Khoa CNTT- ĐHDLVL có thể tham gia khai thác những phần mềm cổng thông tin đã được xây dựng ở nước ta như iPortal, Vportal, VnPortal ứng dụng tích hợp thông tin phục vụ cho một số hoạt động đào tạo ở trường ta và cũng có thể phục vụ cho các quá trình chính phủ điện tử , thương mại điện tử đang diễn ra ở nước ta.

Tài liệu tham khảo :

1. Google 2003. The Anatomy of a large scale search engine.
2. <http://www.google.com>
3. <http://www.panvietnam.com>
4. <http://www.projectliberty.org>
5. <http://java.sun.com>

**Tiến sĩ Nguyễn Lãm
Trưởng khoa Công nghệ Thông tin**