

XÂY DỰNG HỆ THỐNG WEB NGŨ NGHĨA HỖ TRỢ TRA CỨU PHÁP LUẬT VIỆT NAM

A SEMANTIC WEB SYSTEM FOR VIETNAMESE LAWS

Nguyễn Xuân Pha, Huỳnh Công Pháp

Trường Cao đẳng Công nghệ Thông tin, Đại học Đà Nẵng

Email: hcphap@gmail.com

TÓM TẮT

Trong đời sống xã hội, pháp luật có vai trò đặc biệt quan trọng. Tuy nhiên, các hệ thống tra cứu pháp luật hiện tại chưa thật sự hoàn chỉnh và do đó chưa thể đáp ứng tốt yêu cầu của người dùng. Trong bài báo này, chúng tôi đề xuất xây dựng một hệ thống tra cứu và tìm hiểu Pháp luật Việt Nam hoàn chỉnh. Hệ thống mà chúng tôi hướng đến không chỉ dừng lại ở mức độ tìm kiếm và tra cứu mà còn cho phép làm việc như một hệ thống e-learning hỗ trợ sinh viên nghiên cứu, học tập và thi trắc nghiệm Pháp luật Trực tuyến. Hệ thống này cũng sẽ còn là một diễn đàn cho phép những người quan tâm chia sẻ và trao đổi thông tin như một mạng xã hội của những người yêu thích tìm hiểu Pháp luật Việt Nam.

Từ khóa: Web ngữ nghĩa; tra cứu pháp luật; ontology pháp luật; mạng xã hội; học tập trực tuyến

ABSTRACT

Laws play an especially important role in social life. However, the current systems of consulting laws in Vietnam are still incomplete, so they can't meet the needs of users. In this paper, we propose an idea on building a complete system for looking up Vietnamese laws, which supports not only the functions for consulting and search of laws, but also allows itself to act as an e-learning system assisting students to study, learn and take online multiple choice tests of law. This system can also act as a forum supporting those who are interested in sharing and exchanging information concerning Vietnamese laws.

Key words: Semantic web; laws lookup; ontology of laws; social networking; laws e-learning

1. Đặt vấn đề

Trong đời sống xã hội, pháp luật có vai trò đặc biệt quan trọng. Pháp luật dưới góc độ luật học được hiểu như là: Tổng thể các quy tắc xử sự do nhà nước ban hành theo những trình tự thủ tục nhất định và được bảo đảm thực hiện bằng các biện pháp cưỡng chế nhà nước. Nó là phương tiện không thể thiếu bảo đảm cho sự tồn tại, vận hành bình thường của xã hội nói chung và của nền đạo đức nói riêng [1]. Pháp luật không chỉ là một công cụ quản lý Nhà nước hữu hiệu, mà còn tạo môi trường thuận lợi cho sự phát triển của ý thức đạo đức, làm lành mạnh hoá đời sống xã hội và góp phần bồi đắp nên những giá trị mới.

Tuy nhiên, hệ thống pháp luật của nhà nước Việt Nam đưa vào giảng dạy cho công dân còn hạn chế, đặc biệt là đối với tầng lớp học sinh, sinh viên.

Với Internet ngày nay, đó là một kho dữ

liệu khổng lồ nhưng cũng rất hỗn độn, điều này gây không ít khó khăn cho con người khi muốn tìm kiếm thông tin nói chung hay tìm kiếm một vấn đề về pháp luật nào đó nói riêng trên web. Con người thường xuyên gặp phải một lượng lớn thông tin không cần thiết được trả về từ kết quả tìm kiếm, vì thế họ phải tự mình làm nhiệm vụ suy luận, tổng hợp và rút trích thông tin mình cần. Vấn đề đặt ra là làm thế nào để máy tính có thể trợ giúp xử lý tự động thông tin để việc khai thác thông tin trên web đạt hiệu quả hơn.

Hiện nay, ở Việt Nam các website về pháp luật cho phép hỏi đáp những vấn đề về pháp luật nhưng một cách thủ công như (<http://danluat.thuvienphapluat.vn>) hoặc cho phép người sử dụng nhập từ khóa nhưng phải nhập chính xác như: số, chương, điều, khoản,... để tìm kiếm. Như khi tìm kiếm “khung hình phạt tội ăn cắp” sẽ cho kết quả là những website có từ ngữ liên quan “hình phạt khung khiếm cho tội ăn cắp ở Singapore”, “trộm cắp tài sản được bị đơn bãi

nại” nhưng lại không đưa ra được kết quả cụ thể chính xác cần tìm là “Điều 138 Bộ luật Hình sự năm 1999, sửa đổi bổ sung năm 2009” quy định về khung hình phạt.

Bên cạnh những khó khăn và thách thức đó, thì “Web Ngữ nghĩa” ra đời là một công nghệ tối ưu để xây dựng phát triển các hệ thống tìm kiếm. Web Ngữ nghĩa giúp máy tính có thể hiểu được ngữ nghĩa trên web, làm cho việc tìm kiếm, đánh giá, tích hợp thông tin được tiến hành một cách tự động, nhanh chóng và chính xác hơn [2].

Với tầm quan trọng của pháp luật, sự cần thiết về một hệ thống tra cứu pháp luật thông minh và sự ra đời của Web Ngữ nghĩa, chúng tôi đề xuất “Xây dựng hệ thống Web Ngữ nghĩa hỗ trợ tra cứu Pháp luật Việt Nam”.

2. Nghiên cứu tổng quan

2.1. Web Ngữ nghĩa và khả năng ứng dụng trong hỗ trợ tra cứu thông tin

Web Ngữ nghĩa là sự mở rộng của WWW bằng cách thêm vào các mô tả ngữ nghĩa của thông tin dưới dạng mà chương trình máy tính có thể “hiểu” và do vậy cho phép xử lý thông tin hiệu quả hơn [4]. Xét về mặt bản chất, Web Ngữ nghĩa chỉ là một công cụ để con người cũng như máy tính sử dụng để biểu diễn thông tin, hay nói chính xác hơn thì Web Ngữ nghĩa chỉ là một dạng dữ liệu trên web. Khác với các dạng thức dữ liệu được trình bày trong HTML, dữ liệu trong Web Ngữ nghĩa được đánh dấu, phân lớp, mô hình hóa, được bổ sung thêm các thuộc tính, các mối liên hệ,... theo các lĩnh vực cụ thể, qua đó giúp cho các phần mềm máy tính có thể hiểu được dữ liệu và tự động xử lý được những dữ liệu đó.

Có thể kể ra đây những ưu điểm của Web Ngữ nghĩa so với web hiện tại: Máy tính có thể hiểu được thông tin trên web; thông tin được tìm kiếm nhanh chóng và chính xác hơn; khả năng suy luận thông minh; dữ liệu liên kết động.

2.2. Tổng quan hệ thống tra cứu Pháp luật Việt Nam hiện nay

Như trình bày ở phần giới thiệu, hiện nay đang tồn tại một số hệ thống cho phép tra cứu

Pháp luật Việt Nam, đơn cử một số hệ thống tiêu biểu sau đây:

2.2.1. Một số phần mềm tra cứu cài đặt trực tiếp trên máy

- Đầu tiên phải kể đến phần mềm tra cứu văn bản pháp luật của Nhà xuất bản Chính trị Quốc gia – Sự thật đồng sản xuất và phân phối bởi Công ty CP Softech.

Phần mềm tra cứu văn bản Pháp luật cập nhật đầy đủ các văn bản, thông tư, nghị định, nghị quyết, công văn... được nhà nước ban hành đáp ứng nhu cầu của tất cả các lĩnh vực như: Dân sự, Hình sự, Thuế, Lệ phí, Tiền lương, Bảo hiểm xã hội, Bất động sản, Tài chính, Ngân hàng, Xây dựng, Hải quan, Giáo dục, Y tế, Tài nguyên Môi trường, Xuất Nhập khẩu,...

- Một phần mềm tiêu biểu khác đó là phần mềm quản lý và tra cứu văn bản trực tuyến được phát triển bởi Công ty Hỗ trợ Phát triển Phần mềm Luật Việt trụ sở tại: 04 Trần Quang Diệu - Quận Sơn Trà - TP. Đà Nẵng.

Phần mềm này “tự động cập nhật” tất cả các văn bản mới được ban hành từ năm 1945 đến nay trong dữ liệu, gồm:

- Hiến pháp, Luật, Nghị quyết của Quốc hội.
- Pháp lệnh, Nghị quyết của Ủy ban Thường vụ Quốc hội.
- Sắc lệnh, Sắc luật, Lệnh của Chủ tịch nước.
- Các văn bản dưới luật của Chính phủ, Thủ tướng Chính phủ, các Bộ và các Cơ quan ngang Bộ thuộc Chính phủ như: Nghị định, Nghị quyết, Chỉ thị, Quyết định, Thông tư,...
- Các văn bản hướng dẫn.

2.2.2. Hệ thống tra cứu dựa trên nền Web

Bên cạnh những phần mềm cài đặt trực tiếp trên máy tính, chúng ta có thể tìm thấy một số hệ thống tra cứu trên nền web.

- Hệ thống văn bản quy phạm pháp luật được xây dựng bởi bộ Tư pháp nước Cộng hòa Xã hội Chủ nghĩa Việt Nam.

Hệ thống cũng hỗ trợ người dùng tra cứu, nhưng cũng phải dựa vào: tên của văn bản, thông tư, công văn mới cho được kết quả chính xác.

- Trang Thư viện pháp luật:

ThuVienPhapLuat.vn là một trang web chuyên sâu về các vấn đề liên quan tới pháp luật của tất cả các lĩnh vực trong cuộc sống, xã hội với giao diện thân thiện và dễ sử dụng. Trang web giúp mọi người dân có thể tra cứu về một loại văn bản quy phạm pháp luật liên quan đến lĩnh vực mà họ đang quan tâm một cách nhanh chóng và chính xác. Qua trang web, người dân có thể nắm bắt một cách nhanh chóng những thông tin liên quan đến các vấn đề “nóng” của xã hội.

3. Nhận xét chung các hệ thống tra cứu Pháp luật Việt Nam

Phản trên chúng tôi đã giới thiệu một số hệ thống tra cứu Pháp luật Việt Nam hiện nay đang được sử dụng trên thực tế. Nhìn chung các hệ thống này có các ưu điểm, nhược điểm sau.

- Ưu điểm:

Các hệ thống tra cứu Pháp luật Việt Nam có giao diện dễ sử dụng.

Nếu người dùng biết chính xác số hiệu văn bản hoặc tên văn bản thì tìm được kết quả chính xác và nhanh chóng.

Đa số là hệ thống miễn phí cho người dùng (không cần có tài khoản sử dụng hoặc cần tài khoản nhưng không thu phí). Các hệ thống cho tải các văn bản về máy người sử dụng. Hệ thống chạy ổn định.

- Nhược điểm:

Qua nghiên cứu các hệ thống tra cứu Pháp luật Việt Nam hiện nay và tình hình thực tế trong đời sống xã hội, bên cạnh các ưu điểm kể trên thì còn tồn tại rất nhiều hạn chế như:

Chưa hỗ trợ Ngữ nghĩa để hiểu được ngữ nghĩa của từ cần tìm. Ví dụ: Nhập “*Khung hình phạt tội ăn trộm*” thì không cho ra được kết quả là: “*Điều 138 Bộ luật Hình sự năm 1999, sửa đổi bổ sung năm 2009*”.

Chưa tạo ra môi trường hỗ trợ việc tìm hiểu cũng như học tập trực tuyến về pháp luật (E-learning).

Chưa cho phép người dùng tham gia đóng góp, chia sẻ, thảo luận những vấn đề về pháp luật quan tâm (social network).

Cơ sở dữ liệu không được cập nhật một

cách tự động. Không trích rút được từ các hệ thống khác.

Không hỗ trợ một số tính năng trực tuyến, như không trả lời tự động mà phải thông qua luật sư cho nên phải có thời gian chờ đợi,...

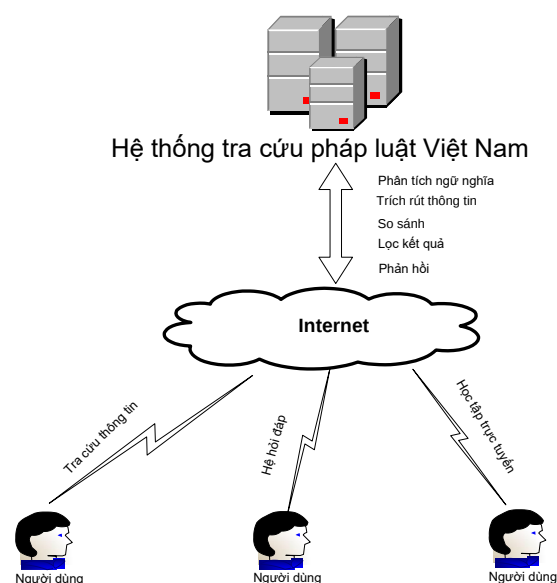
4. Đề xuất hệ thống tra cứu Pháp luật Việt Nam hoàn chỉnh

4.1. Đặt tử tổng quan

Như trình bày ở mục trên, các hệ thống tra cứu pháp luật hiện nay còn hạn chế, chưa hỗ trợ tốt nhu cầu của người dùng. Vì vậy, một hệ thống tra cứu pháp luật hoàn chỉnh cần khắc phục được các yếu tố trên và có những chức năng hỗ trợ sau:

- Đầu tiên hệ thống phải lọc và trích rút từ mạng internet các dữ liệu liên quan đến pháp luật một cách tự động, nhanh chóng để đáp ứng các nhu cầu của người dùng.

- Hệ thống cho phép tra cứu thông tin pháp luật (công văn, văn bản, thông tư,...) dựa vào ngữ nghĩa, hệ thống phải hiểu được ngữ nghĩa của từ cần tìm để cho kết quả chính xác nhất.



Hình 1. Mô hình tổng quan về một hệ thống tra cứu pháp luật

- Là một hệ hỏi đáp tự động tức thời, trả lời các vấn đề thông thường nhất của người dùng về pháp luật một cách tự động nhanh chóng.

- Là một hệ học tập trực tuyến (E-learning), hỗ trợ người dùng học tập các kiến thức về Pháp luật Việt Nam thông qua mạng Internet.

4.2. Đề xuất quy trình xây dựng hệ thống tra cứu Pháp luật Việt Nam

Đối với bài toán xây dựng hệ thống tra cứu Pháp luật Việt Nam, chúng tôi đề xuất xây dựng chương trình trên nền web cụ thể Web Ngữ nghĩa, công việc là một quy trình gồm các bước sau:

- Thứ nhất, tiến hành nghiên cứu những gì liên quan đến Pháp luật Việt Nam.
- Thứ hai, tìm hiểu các hệ thống tra cứu Pháp luật Việt Nam hiện nay.
- Thứ ba, tìm hiểu lý thuyết liên quan đến Web Ngữ nghĩa.
- Thứ tư, xây dựng Ontology về Pháp luật Việt Nam bằng công cụ Protégé.
- Thứ năm, sử dụng các công cụ như: Semweb, OwlDotNetApi, Hệ truy vấn SPARQL,... để xây dựng hệ thống tra cứu theo hướng Web Ngữ nghĩa.
- Cuối cùng, xây dựng các chức năng như: Hệ hỏi đáp, hệ học tập trực tuyến,... vào hệ thống để hoàn thành một hệ thống đa nhiệm.

Xây dựng Ontology về pháp luật

Đối với bài toán xây dựng hệ thống tra cứu Pháp luật Việt Nam thì vấn đề tìm kiếm dữ liệu ngữ nghĩa ở đây là dữ liệu về pháp luật. Vì vậy tất cả các lĩnh vực về pháp luật phải được lưu trong một lớp, và đây là lớp quan trọng nhất của bài toán cần xây dựng.

Dữ liệu liên quan về pháp luật như: các văn bản, các quy định, các thông tư, các vụ án, các câu hỏi liên quan pháp luật,... Mỗi điều kiện tìm kiếm ta xây dựng nên những lớp thông tin tương ứng.

Khi xây dựng cơ sở dữ liệu ngữ nghĩa về pháp luật việc lưu trữ các thông tin dưới dạng các lớp (các class) với các khai báo thông tin nên sau khi xây dựng ta có thể bổ sung trong quá trình sử dụng. Còn dữ liệu ngữ nghĩa về thông tin pháp luật là các lớp con của lớp pháp luật.

4.2.1. Mô tả các lớp

Trong phần này chúng tôi đưa ra một số lớp quan trọng trong vấn đề tra cứu pháp luật, các lớp khác sẽ được thêm vào bằng phần mềm Protégé hoặc trong quá trình sử dụng sau này.

- Vanbanphapluat: là tất cả các văn bản

pháp luật, lớp này có các lớp con như: Boluat; luat; congvan; chithi; lenh; thongtu; nghidinh; nghiquyet;...

- Loaitoidanh: là các loại tội danh, lớp này có các lớp con như: moitruong; matuy; XPanninhquocgia; XPsohuu; XPhonnhangiadin; ...

- Loaitoipham: là các loại tội phạm lớp này có các lớp con như: dacbietnghiemtrong; ratnghiemtrong; nghiemtrong; itnghiemtrong.

- Linhvuc: là danh mục tất cả các cá thể của pháp luật mà ta lưu trữ phục vụ tìm kiếm sau này. Việc định nghĩa các lĩnh vực pháp luật và các thông tin liên quan đến nó sẽ là nguồn dữ liệu thông minh để tìm kiếm nhanh và cho kết quả chính xác. Đây là lớp chính chứa thông tin về pháp luật có liên quan đến các cá thể của các lớp thông tin.

4.2.2. Mô tả các thuộc tính

Các thuộc tính của một lớp biểu thị quan hệ của nó với một lớp. Thông thường các thuộc tính được quy định kiểu dữ liệu để hỗ trợ việc xây dựng chương trình có tính ràng buộc hơn.

Trong Web Ngữ nghĩa việc lưu trữ dữ liệu thông qua XML nên có khả năng tùy biến rất cao, một trong những khả năng tùy biến đó là sử dụng thuộc tính một cách linh động.

4.2.3. Các cá thể

Mục đích cá thể là dùng để lưu trữ dữ liệu riêng biệt của đối tượng. Dữ liệu của các cá thể là các dữ liệu đơn và dữ liệu này được lưu trong một trường thông tin.

5. Xây dựng chương trình và kết quả nghiên cứu

5.1. Các công cụ cần thiết để xây dựng Web Ngữ nghĩa

Dưới đây là các công cụ cần thiết để xây dựng Web Ngữ nghĩa:

5.1.1. Công cụ xây dựng Ontology Protégé

Protégé cho phép người dùng sử dụng tạo ra các Ontology để phát triển Web Ngữ nghĩa theo đúng chuẩn của ngôn ngữ W3C OWL [3].

5.1.2. Hệ truy vấn SPARQL

SPARQL là một ngôn ngữ để truy cập

thông tin từ các đồ thị RDF, nó cung cấp các tính năng trích thông tin trong các dạng của URI.

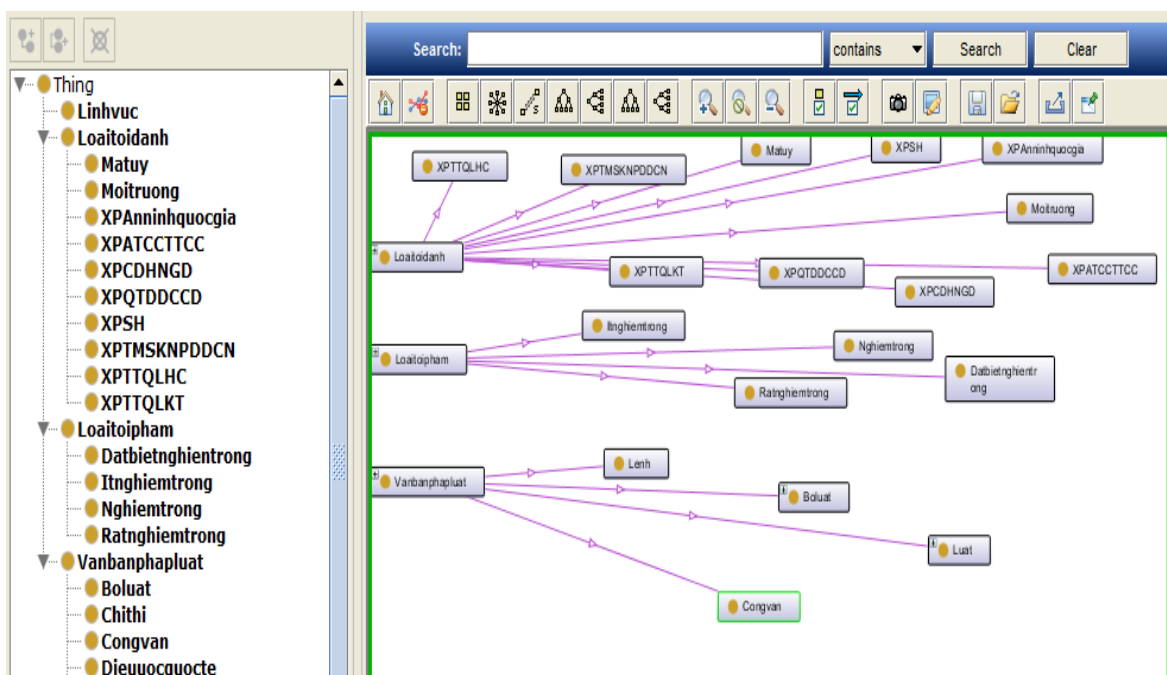
5.1.3. Công cụ SemWeb

SemWeb cung cấp các tính năng cốt lõi như đọc/ ghi dữ liệu XML với bộ ba RDF, liên tục lưu trữ dữ liệu với nền tảng SQL và các truy vấn SPARQL cơ bản.

5.2. Kết quả nghiên cứu

Dưới đây là một số kết quả mà nhóm

chúng tôi bước đầu cơ bản thu được. Mặc dù, kết quả này còn khiêm tốn so với tham vọng cũng như mục tiêu đặt ra, nhưng những kết quả đạt được là rất quan trọng, sẽ là tiền đề để chúng tôi xây dựng được một hệ thống tra cứu Pháp luật Việt Nam hoàn chỉnh. Kết quả nghiên cứu quan trọng nhất phải kể đến đó là chúng tôi xây dựng được Ontology về Pháp luật Việt Nam (Hình 2) và chức năng tra cứu ngữ nghĩa về Pháp luật Việt Nam (Hình 3)



Hình 2. Các lớp trong Ontology

6. Kết luận

Trong bài báo này, chúng tôi đã đề xuất một mô hình hệ thống tra cứu Pháp luật Việt Nam trên cơ sở ứng dụng Web Ngữ nghĩa, bao gồm nhiều chức năng tra cứu cơ bản và nâng cao. Ý tưởng một hệ thống tra cứu Pháp luật mà chúng tôi hướng đến không dừng lại ở mức độ tìm kiếm và tra cứu mà còn cho phép làm việc như một hệ thống e-learning hỗ trợ sinh viên nghiên cứu và thi trắc nghiệm Pháp luật Trực tuyến. Hệ thống này cũng sẽ là một kênh thông

tin để chia sẻ và trao đổi của những người muốn tìm hiểu Pháp luật Việt Nam.

Mặc dù, ý tưởng về một hệ thống tối ưu như vậy nhưng hệ thống thực nghiệm xây dựng được chỉ mới bước đầu cho phép tra cứu và tìm kiếm theo ngữ nghĩa. Do vậy, vấn đề trước mắt và lâu dài của nhóm chúng tôi là sẽ tiếp tục hiện thực ý tưởng trên để cộng đồng người Việt Nam chúng ta có một hệ thống hỗ trợ tra cứu và tìm hiểu Pháp luật Việt Nam tốt nhất.



Hình 3. Giao diện tra cứu Pháp luật Việt Nam

TÀI LIỆU THAM KHẢO

- [1] Lê Thị Tuyết Ba. Vai trò của pháp luật đối với việc hình thành và phát triển ý thức đạo đức ở nước ta hiện nay, *Tạp chí Triết học*, 2008.
- [2] Huỳnh Công Pháp, Ngô Thị Thanh Trang, Giải pháp ứng dụng Web Ngữ nghĩa vào việc tổ chức và tra cứu hiệu quả thông tin môn học, ngành học trong các trường Đại học, *Tạp chí Khoa học và Công nghệ, Đại học Đà Nẵng*, số 9(58).quyển 3, 2012.
- [3] Nguyễn Việt Cường, *Semantic Web tương lai của Word Wide Web*, 2004
- [4] H. Peter Alesso, Craig F. Smith, Thinking on the Web: Berners-Lee, Godel and Turing, John Wiley & Sons Inc, 2006
- [5] Berners-Lee, T., Hendler, J., and Lassila, O., The Semantic Web, Scientific American, 2001.

(BBT nhận bài: 23/09/2013, phản biện xong: 01/11/2013)