1. Ngôn ngữ sử dụng trong chương trình này là tiếng việt.
2. Để chạy chương trình ta chạy file *index.jsp*.
3. Đầu tiên, chương trình sẽ vào màn hình giao diện *giới thiệu* (hình 1). Giao diện này sẽ giới thiệu mục đích làm đề tài và nêu tên các thành viên trong nhóm. Nó có đường link để thực thi quá trình làm giàu Ontology.



Hình 1: giới thiệu

1. Khi ta nhấn vào nút Nhấn vào đây để chạy chương trình, nó sẽ sang giao diện màn hình *Nhập dữ liệu tìm kiếm* (hình 2). Giao diện này được chia làm ba cột chính: cột hình, cột Ontology và cột Option.
   1. Trước tiên, ta cần vào cột Ontology, click vào những khái niệm mà ta muốn làm giàu. Tên những khái niệm này sẽ hiện ra bên phần *Những khái niệm được chọn*. Ta có thể làm giàu nhiều lớp một lúc, nhưng tốt đa chỉ được bốn lớp. Ở đây, ta chọn khái niệm *Phần mềm*.
   2. Bên phần *Chọn công cụ tìm kiếm*, ta cần check vào checkbox *Google* hay *Yahoo* và chọn số lượng link cần tìm ở bên cạnh.
   3. Sau khi đã chọn lựa xong, ta nhấn nút *Tìm kiếm*.



Hình 2: Nhập dữ liệu tìm kiếm

1. Chương trình bây giờ sẽ ở giao diện màn hình *Danh sách các links* (hình 3). Giao diện này chia làm hai phần. Giao diện bên phải sẽ hiện ra tất các link (và tiêu đề tương ứng) mà công cụ tìm kiếm tra được. Phần bên trái sẽ thể hiện các giai đoạn chạy. Có tất cả bốn giai đoạn: Danh sách các link, Tải về máy và phân lớp, Rút trích tài liệu công nghệ thông tin, Cập nhật vào Ontology. Các giai đoạn này tuần tự từ trên xuống dưới và không thể quay ngược lại. Lúc này ta đang ở giai đoạn *Danh sách các link*, giai đoạn tiếp theo là *Tải về máy và phân lớp*. Muốn đến giai đoạn tiếp theo, ta chỉ việc nhấn vào link *Tải về máy và phân lớp*.



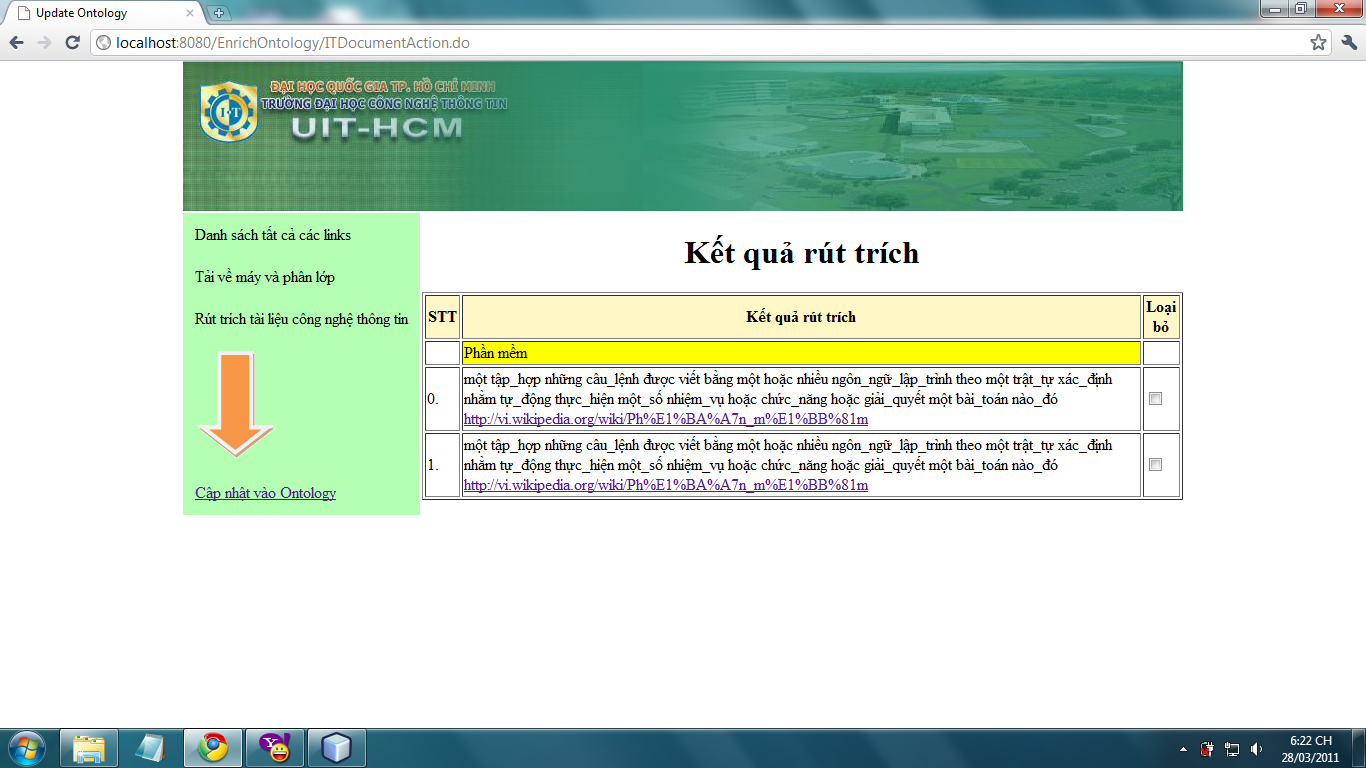
Hình 3: Danh sách các links

1. Giai đoạn *Tải về máy và phân lớp* (tương ứng là giao diện màn hình *Phân lớp* (hình 4)) sẽ thực hiện các công đoạn: tải dữ liệu về máy theo các đường link đã có, tách từ, loại bỏ stopword, phân lớp theo SVM. Khi giai đoạn kết thúc thì chương trình sẽ hiện ra danh sách các link (và tiêu đề tương ứng) được chọn ở phần bên phải. Còn phần bên trái sẽ hiện thị cho biết ta đang ở giai đoạn *Tải về máy và phân lớp* và giai đoạn tiếp theo là *Rút trích tài liệu công nghệ thông tin*.



Hình 4: Phân lớp

1. Giai đoạn *Rút trích tài liệu công nghệ thông tin* (tương ứng là giao diện màn hình *Rút trích* (hình 5)) sẽ thực hiện công đoạn rút trích ra định nghĩa từ những link được chọn. Khi giai đoạn kết thúc thì chương trình sẽ hiện ra danh sách các đoạn được rút ra và nguồn của chúng ở phần bên phải. Người dùng có thể check vào các ô checkox tương ứng để loại bỏ những đoạn không liên quan đến khái niệm cần làm giàu. Còn phần bên trái sẽ hiện thị cho biết ta đang ở giai đoạn *Rút trích tài liệu công nghệ thông tin* và giai đoạn tiếp theo là *Cập nhật Ontology*.



Hình 5: Rút trích

1. Giai đoạn *Cập nhật Ontology* sẽ thực hiện công đoạn: cập nhật các định nghĩa và nguồn tương ứng với nó vào Ontology. Sau đó, chương trình sẽ hiện lên thông báo về xem giai đoạn cập nhật Ontology có thành công hay không (hình 6). Đến đây, ta đã kết thúc toàn bộ quá trình làm giàu Ontology. Nếu ta muốn tiếp tục làm giàu một khái niệm nào đó thì ta nhấn nút *Trở lại trang đầu*.



Hình 6:Kết quả

Thực nghiệm chương trình:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| stt | Tên từ khóa | Số từ khóa | Số link chọn | Thời gian chạy chương trình | Số định nghĩa thu được | Số định nghĩa đúng (không tính trùng nhau) |
| 1 | Phần mềm | 1 | 10 | 1 phút 35 giây | 2 | 1 |
| 2 | Tin học | 1 | 10 | 1 phút 17 giây | 2 | 1 |
| 3 | Tin học | 1 | 25 | 2 phút 50 giây | 6 | 1 |
| 4 | Phần mềm | 1 | 30 | 1 phút 55 giây | 12 | 1 |
| 4 | Phần mềm – Tin học | 2 | 10 | 1 phút 48 giây | 4 | 2 |
| 5 | Phần mềm – Tin học | 2 | 20 | 5 phút 14 giây | 13 | 2 |
| 6 | Công nghệ thông tin | 1 | 10 | 30 giây | 4 | 2 |
| 7 | Hệ thống thông tin | 1 | 10 | 55 giây | 0 | 0 |
| 8 | Hợp đồng | 1 | 15 | 1phút 52 giây | 2 | 1 |
| 9 | Bộ nhớ ảo – Hệ điều hành | 2 | 60 | 10 phủ 16 giây | 0 | 0 |
| 10 | Phần mềm – Tin học – Hệ điều hành | 3 | 30 | 3 phút 39 giây | 11 | 4 |
| 11 | Phần mềm – Tin học – Hệ điều hành – Ngôn ngữ lập trình | 4 | 20 | 6 phút 38 giây | 11 | 4 |
| 12 | Lập trình | 1 | 50 | 12 phút 5 giây | 1 | 0 |

Đánh giá chương trình:

* Phần làm được: chương trình có thể làm giàu theo cách thêm định nghĩa cho những khái niệm trong hệ thống Ontology chuyên ngành công nghệ thông tin tiếng việt.
* Phần chưa làm:
  + Chưa hoàn thiện phần rút trích.
  + Thời gian chạy chương trình còn lâu.
  + Chưa làm giàu quan hệ cho hệ thống Ontology.

THỬ NGHIỆM

* 1. Cài đặt:
* Hệ điều hành WinXP/Vista/Win7.
* Cài đặt phần mềm NetBeans 6.7.
* Cài đặt jdk 1.6, jre6.
* Cần có mạng.
* Cần chép file resource ra ổ đĩa E:/.

1. Ngôn ngữ sử dụng trong chương trình này là tiếng việt.
2. Để chạy chương trình ta chạy file *index.jsp*.
3. Đầu tiên, chương trình sẽ vào màn hình giao diện *giới thiệu* (hình 1). Giao diện này sẽ giới thiệu mục đích làm đề tài và nêu tên các thành viên trong nhóm. Nó có đường link để thực thi quá trình làm giàu Ontology.



Hình 1: giới thiệu

1. Khi ta nhấn vào nút Nhấn vào đây để chạy chương trình, nó sẽ sang giao diện màn hình *Nhập dữ liệu tìm kiếm* (hình 2). Giao diện này được chia làm ba cột chính: cột hình, cột Ontology và cột Option.
   1. Trước tiên, ta cần vào cột Ontology, click vào những khái niệm mà ta muốn làm giàu. Tên những khái niệm này sẽ hiện ra bên phần *Những khái niệm được chọn*. Ta có thể làm giàu nhiều lớp một lúc, nhưng tốt đa chỉ được bốn lớp. Ở đây, ta chọn khái niệm *Phần mềm*.
   2. Bên phần *Chọn công cụ tìm kiếm*, ta cần check vào checkbox *Google* hay *Yahoo* và chọn số lượng link cần tìm ở bên cạnh.
   3. Sau khi đã chọn lựa xong, ta nhấn nút *Tìm kiếm*.



Hình 2: Nhập dữ liệu tìm kiếm

1. Chương trình bây giờ sẽ ở giao diện màn hình *Danh sách các links* (hình 3). Giao diện này chia làm hai phần. Giao diện bên phải sẽ hiện ra tất các link (và tiêu đề tương ứng) mà công cụ tìm kiếm tra được. Phần bên trái sẽ thể hiện các giai đoạn chạy. Có tất cả bốn giai đoạn: Danh sách các link, Tải về máy và phân lớp, Rút trích tài liệu công nghệ thông tin, Cập nhật vào Ontology. Các giai đoạn này tuần tự từ trên xuống dưới và không thể quay ngược lại. Lúc này ta đang ở giai đoạn *Danh sách các link*, giai đoạn tiếp theo là *Tải về máy và phân lớp*. Muốn đến giai đoạn tiếp theo, ta chỉ việc nhấn vào link *Tải về máy và phân lớp*.



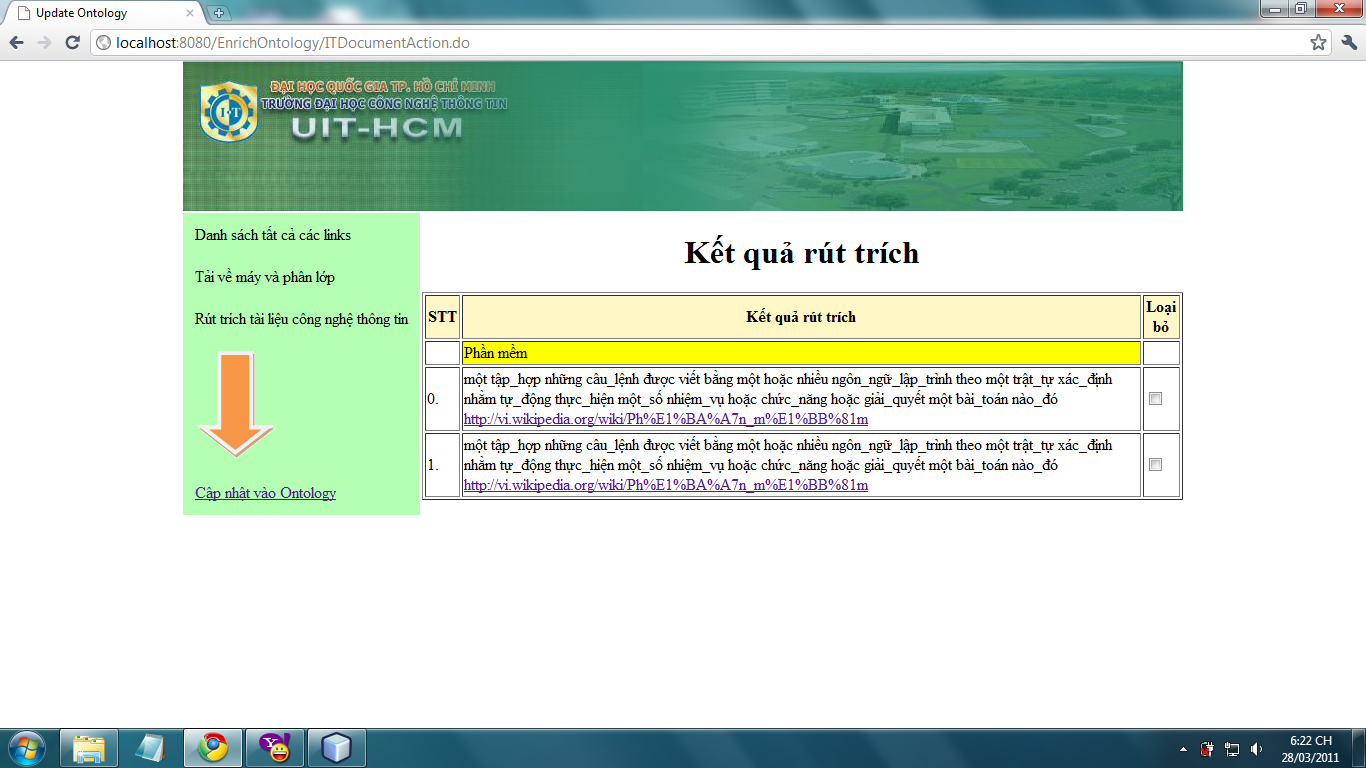
Hình 3: Danh sách các links

1. Giai đoạn *Tải về máy và phân lớp* (tương ứng là giao diện màn hình *Phân lớp* (hình 4)) sẽ thực hiện các công đoạn: tải dữ liệu về máy theo các đường link đã có, tách từ, loại bỏ stopword, phân lớp theo SVM. Khi giai đoạn kết thúc thì chương trình sẽ hiện ra danh sách các link (và tiêu đề tương ứng) được chọn ở phần bên phải. Còn phần bên trái sẽ hiện thị cho biết ta đang ở giai đoạn *Tải về máy và phân lớp* và giai đoạn tiếp theo là *Rút trích tài liệu công nghệ thông tin*.



Hình 4: Phân lớp

1. Giai đoạn *Rút trích tài liệu công nghệ thông tin* (tương ứng là giao diện màn hình *Rút trích* (hình 5)) sẽ thực hiện công đoạn rút trích ra định nghĩa từ những link được chọn. Khi giai đoạn kết thúc thì chương trình sẽ hiện ra danh sách các đoạn được rút ra và nguồn của chúng ở phần bên phải. Người dùng có thể check vào các ô checkox tương ứng để loại bỏ những đoạn không liên quan đến khái niệm cần làm giàu. Còn phần bên trái sẽ hiện thị cho biết ta đang ở giai đoạn *Rút trích tài liệu công nghệ thông tin* và giai đoạn tiếp theo là *Cập nhật Ontology*.



Hình 5: Rút trích

1. Giai đoạn *Cập nhật Ontology* sẽ thực hiện công đoạn: cập nhật các định nghĩa và nguồn tương ứng với nó vào Ontology. Sau đó, chương trình sẽ hiện lên thông báo về xem giai đoạn cập nhật Ontology có thành công hay không (hình 6). Đến đây, ta đã kết thúc toàn bộ quá trình làm giàu Ontology. Nếu ta muốn tiếp tục làm giàu một khái niệm nào đó thì ta nhấn nút *Trở lại trang đầu*.



Hình 6:Kết quả

Thực nghiệm chương trình:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| stt | Tên từ khóa | Số từ khóa | Số link chọn | Thời gian chạy chương trình | Số định nghĩa thu được | Số định nghĩa đúng (không tính trùng nhau) |
| 1 | Phần mềm | 1 | 10 | 1 phút 35 giây | 2 | 1 |
| 2 | Tin học | 1 | 10 | 1 phút 17 giây | 2 | 1 |
| 3 | Tin học | 1 | 25 | 2 phút 50 giây | 6 | 1 |
| 4 | Phần mềm | 1 | 30 | 1 phút 55 giây | 12 | 1 |
| 4 | Phần mềm – Tin học | 2 | 10 | 1 phút 48 giây | 4 | 2 |
| 5 | Phần mềm – Tin học | 2 | 20 | 5 phút 14 giây | 13 | 2 |
| 6 | Công nghệ thông tin | 1 | 10 | 30 giây | 4 | 2 |
| 7 | Hệ thống thông tin | 1 | 10 | 55 giây | 0 | 0 |
| 8 | Hợp đồng | 1 | 15 | 1phút 52 giây | 2 | 1 |
| 9 | Bộ nhớ ảo – Hệ điều hành | 2 | 60 | 10 phủ 16 giây | 0 | 0 |
| 10 | Phần mềm – Tin học – Hệ điều hành | 3 | 30 | 3 phút 39 giây | 11 | 4 |
| 11 | Phần mềm – Tin học – Hệ điều hành – Ngôn ngữ lập trình | 4 | 20 | 6 phút 38 giây | 11 | 4 |
| 12 | Lập trình | 1 | 50 | 12 phút 5 giây | 1 | 0 |

Đánh giá chương trình:

* Phần làm được: chương trình có thể làm giàu theo cách thêm định nghĩa cho những khái niệm trong hệ thống Ontology chuyên ngành công nghệ thông tin tiếng việt.
* Phần chưa làm:
  + Chưa hoàn thiện phần rút trích.
  + Thời gian chạy chương trình còn lâu.
  + Chưa làm giàu quan hệ cho hệ thống Ontology.

Cài đặt:

* Hệ điều hành WinXP/Vista/Win7.
* Cài đặt phần mềm NetBeans 6.7.
* Cài đặt jdk 1.6, jre6.
* Cần có mạng.
* Cần chép file resource ra ổ đĩa E:/.