

## Mục lục

|  |    |
|--|----|
| Chuẩn chú thích tài liệu.....                                    | 1  |
| Các Entity (thực thể).....                                       | 1  |
| Product (sản phẩm).....  | 1  |
| Organization (tổ chức).....                                      | 2  |
| Person (con người).....  | 2  |
| Location (địa điểm).....   | 2  |
| Đồng tham chiếu.....   | 2  |
| Feature (tính năng).....   | 2  |
| Phân loại theo cách diễn đạt.....                                | 2  |
| Các feature lồng nhau.....                                       | 3  |
| Opinion expression (diễn đạt ý kiến).....                        | 3  |
| Chủ quan hay khách quan?.....                                    | 3  |
| Tích cực hay tiêu cực?.....                                      | 3  |
| Trực tiếp và gián tiếp.....                                      | 3  |
| Các thuộc tính.....  | 4  |
| Valence shifter (từ chuyển nghĩa).....                           | 4  |
| Yêu cầu.....   | 4  |
| Xây dựng ngữ liệu.....   | 4  |
| Cộng tác viên.....   | 4  |
| ePi-Teamware: Công cụ cộng tác và quản lý xây dựng ngữ liệu..... | 5  |
| Chức năng.....   | 5  |
| Thiết kế.....  | 7  |
| Mã nguồn.....  | 7  |
| Triển khai (deploy).....   | 7  |
| ReferenceSchemaAnnotationEditor: Công cụ chú thích.....          | 7  |
| Nạp plugin.....  | 7  |
| Chú thích các thực thể và tính năng.....                         | 7  |
| Chú thích các tính năng, ý kiến và từ chuyển nghĩa.....          | 8  |
| Kinh nghiệm chú thích hiệu quả.....                              | 11 |
| Công cụ so sánh.....   | 12 |

## Chuẩn chú thích tài liệu

### Các Entity (thực thể)

Ở đây ta quan tâm đến các **thực thể có tên**. Chúng là các sản phẩm, thương hiệu, tổ chức, nhân vật,... Khi chú thích thực thể như vậy ta **chỉ đánh dấu tên** mà bỏ qua những từ đi kèm.

VD: bộ vi xử lý Intel Core 2 Duo T7500, Model Asus S6Fm, hệ điều hành Linux

Thuộc tính **co-reference** = Đồng tham chiếu (đề cập đến một thực thể đã được nhắc đến trước đó).

#### *Product (sản phẩm)*

Trường hợp tên sản phẩm chứa tên nhà sản xuất thì **chú thích cả cụm như là tên sản phẩm** mà không chú thích tên nhà sản xuất nữa.

VD: Loạt sản phẩm mới như Asus Lamborghini VX2 (Product), Asus S6Fm (Product)... đã đem về niềm tự hào cho thương hiệu này trên khắp thế giới.

Đôi khi người ta dùng tên hãng để chỉ một nhóm thực thể sản xuất bởi hãng đó, trường hợp này ta chú thích Product (KHÔNG dùng thẻ Organization) vào tên hãng.

VD: Laptop ASUS (Product) được thiết kế mỏng và nhẹ, và ổn định nhờ có bo mạch chủ ASUS (Product) bên trong, hệ thống làm mát ATDT2 (Product) độc quyền, sự cơ động và sự bảo mật an toàn.

Organization (tổ chức)

VD: cty Sony, Hội Liên hiệp Khoa học Việt Nam, trường đại học Bách Khoa Hà Nội,...

Person (con người)

VD: chủ tịch Hồ Chí Minh, danh ca Whitney Houston,...

Gender = { male, female }

Location (địa điểm)

VD: Việt Nam, Đức, Ireland, thủ đô Hà Nội, tp. HCM...

Đồng tham chiếu

Bài viết có thể nhắc đến một thực thể rất nhiều lần, sử dụng tên riêng của nó hoặc không, ta gọi hiện tượng này là đồng tham chiếu. Có thể phân ra thành những loại chính:

- Dùng tên: ASUS UIF = UIF, Eee Pad Slider = Eee Pad = Slider, Motorola = Moto, Volkswagen = Con bọ,...
- Đại từ: Chiếc máy này,...
- Khác: Máy, xe, hãng, nhà sản xuất,...

## Feature (tính năng)

Các chức năng, bộ phận, tính chất của một thực thể được gọi chung là feature (tính năng). VD:

- **Chức năng**: soạn thảo văn bản, chụp ảnh, đàm thoại video,...
- **Bộ phận**: bàn phím, nút bấm, vỏ, dây,...
- **Tính chất**: trọng lượng, hình dạng, sự linh hoạt,...

Các thuộc tính của thẻ Feature:

- feature-of = tham chiếu đến feature / entity
- co-reference = tham chiếu đến feature khác

Cần đánh dấu một và chỉ một trong hai thuộc tính trên.

Phân loại theo cách diễn đạt

- **Trực tiếp**: gọi tên một bộ phận hoặc một thuộc tính nào đó.
  - bàn phím, loa, độ sắc nét
- **Gián tiếp**: đề cập đến một bộ phận hoặc (thường là) một thuộc tính nào đó. Feature dạng này có thể là động từ
  - Điện thoại này trông rất tinh tế. - “trông” là feature đề cập đến vẻ ngoài
- **Ấn**: không xuất hiện trong câu nhưng được đề cập thông qua opinion expression
  - Điện thoại này bị lỗi hồng tâm - “lỗi hồng tâm” là opinion expression đề cập đến feature “camera”.

Các feature lồng nhau

Với thiết kế độc đáo của bàn phím... - “thiết kế” là feature con của “bàn phím”, opinion expression là “độc đáo” đề cập đến “thiết kế”.

### Opinion expression (diễn đạt ý kiến)

Nói lên một ý kiến đánh giá chủ quan của người viết/nói.

Mỗi một ý kiến có một polarity (hướng), rơi vào một trong 3 trường hợp:

- Positive (tích cực): nhiều, nhanh, tốt, rẻ, đặc biệt,...
- Negative (tiêu cực): xấu, tệ, hỏng, dở, lỗi,...
- Neutral (trung tính): bình thường,...

Chủ quan hay khách quan?

Khi chú thích ta chỉ quan tâm đến các ý kiến đánh giá chủ quan của người viết/nói mà bỏ qua các thông tin khách quan.

Các ý kiến khách quan thường chứa số liệu hoặc chỉ mô tả mà không đánh giá. VD: "Để chế tạo một chiếc giày này cần qua 10 công đoạn, công đoạn đầu tiên là..."

Một dấu hiệu đơn giản để nhận biết một ý kiến chủ quan là khi không có tiêu chuẩn / quy định / văn bản nào hỗ trợ ý kiến đó. Chẳng hạn: "UIF mỏng và nhẹ" - không có văn bản nào quy định độ dày bao nhiêu thì được gọi là mỏng hay trọng lượng bao nhiêu có thể coi là nhẹ.

Ý kiến chủ quan cũng được thể hiện gián tiếp qua các từ đệm (cả, những, chỉ,...) và cách chọn từ (nhìn >< chiêm ngưỡng, có khả năng >< được ban cho khả năng,...)

Vài trường hợp hay gặp:

| Khách quan                                       | Chủ quan  |
|--|---|
| <i>Chiếc máy nặng 1.950gr</i>                    | <i>Chiếc máy khá <u>nhẹ</u></i>                                 |
| <i>Máy có thể chụp trong đêm</i>                 | <i>Máy thể chụp <u>cả</u> trong đêm</i>                         |
| <i>... để sử dụng máy, bạn trượt bàn phím...</i> | <i>... để sử dụng máy, bạn <u>chỉ</u> cần trượt bàn phím...</i> |
| <i>A giống B</i>                                 | <i>A <u>bất ngờ</u> / <u>nhái</u> B</i>                         |
| <i>nhìn vào chiếc máy,...</i>                    | <i><u>chiêm ngưỡng</u> chiếc máy,...</i>                        |

Tích cực hay tiêu cực?

Sau khi xác định được ý kiến chủ quan ta cần đánh giá polarity (hướng) của ý kiến. Việc này đôi khi phụ thuộc vào người chú thích nhưng ta cần cố gắng đánh giá dựa vào mục đích, công dụng của đối tượng được nói đến hoặc theo suy nghĩ thường gặp của mọi người.

Trực tiếp và gián tiếp

Ý kiến trực tiếp đánh giá về sản phẩm, ý kiến gián tiếp đánh giá tác động của sản phẩm lên người sử dụng, ví dụ: "Sở hữu chiếc máy ảnh này, bạn có thể thoải mái với chuyến du lịch của mình mà vẫn không bỏ lỡ những khung cảnh góc rộng tuyệt đẹp."

Đối với ý kiến gián tiếp ta chú thích đối tượng chịu tác động của sản phẩm như là feature của sản phẩm đó.

VD1: Sở hữu chiếc máy ảnh này (Product), bạn có thể thoải mái (Opinion expression của Feature 1) với chuyến du lịch (Feature 1) của mình mà vẫn không (Valence shifter) bỏ lỡ (Opinion Expression của

Feature 2) những khung cảnh (Feature 2) góc rộng tuyệt đẹp.

VD2: Có lẽ (Valence shifter) không ai (Valence shifter) có thể (Opinion expression của Feature 2) dùng một tay để trượt (Feature 2 con của Feature 1) bàn phím này (Feature 1) ra được.

Các thuộc tính

- polarity = { POSITIVE, NEGATIVE, NEUTRAL }
- opinion-of = Tham chiếu đến feature / entity

### Valence shifter (từ chuyển nghĩa)

Là những từ, cụm từ ảnh hưởng đến ý nghĩa của ý kiến trong câu theo một trong ba cách sau:

- Đảo ngược (negation): *không, không ai, không bao giờ,...*
- Nhân mạnh (intensifier): *rất, hoàn toàn, tuyệt đối,...*
- Giảm nhẹ (diminisher): *hơi, có vẻ, trông như,...*

direction = { NEGATION, INTENSIFIER, DISMINISHER }

shifter-of = Tham chiếu đến opinion word (nếu có)

### Yêu cầu

- **Entity**: Chú thích **tất cả** các thực thể có tên xuất hiện trong văn bản và các từ đồng tham chiếu đến chúng.
- **Feature**: Chỉ chú **tất cả** các feature xuất hiện gắn với thực thể nhất định, bao gồm **tất cả các loại thực thể**, chọn từ, cụm từ **ngắn nhất** đủ xác định feature được nói đến.
- **Opinion expression**: Chú thích cụm từ **ngắn nhất** đủ diễn đạt ý kiến **chủ quan của người viết** đánh giá sản phẩm là tiêu cực hay tích cực (không bao trùm valence shifter).
- **Valence shifter**: Đánh dấu **nếu có ý kiến đánh giá**.
- Chú thích cả các **ý kiến gián tiếp**

## Xây dựng ngữ liệu

### Cộng tác viên

(Tất cả CTV đều là SV Đại học Bách Khoa HN)

| STT | Họ và tên        | SN       | Giới tính | Chuyên ngành |
|-----|------------------|----------|-----------|--------------|
| 1   | Nguyễn Văn Anh   |          |           |              |
| 2   | Đông Như Hà      | 01/15/90 | Nam       | SP kỹ thuật  |
| 3   | Phạm Mỹ Hào      | 11/27/91 | Nữ        | CNTT         |
| 4   | Nguyễn Phi Hiệp  | 4/30/93  | Nam       |              |
| 5   | Bùi Duy Khánh    |          | Nam       |              |
| 6   | Nguyễn Hồng Quân |          |           |              |
| 7   | Phạm Quốc Thắng  |          |           |              |
| 8   | Lưu Thanh Tùng   |          | Nam       | CNTT         |

|    |                  |         |     |                 |
|----|------------------|---------|-----|-----------------|
| 9  | Nguyễn Minh Tùng | 4/11/90 | Nam | Kinh tế quản lý |
| 10 | Đặng Đức Tú      |         | Nam | Toán tin        |

### Quy trình xây dựng ngữ liệu

Đội CTV được chia làm hai nhóm chú thích song song cùng một lượng tài liệu. Như vậy mỗi tài liệu sẽ có hai bản chú thích để có thể tính độ đồng thuận. Đây là một tiêu chí quan trọng để đánh giá chất lượng của một bộ ngữ liệu xây dựng bằng tay.

Để chú thích một tài liệu, trước hết CTV phải nhận chú thích tài liệu thông qua trang web của dự án.

Khi một CTV đã nhận tài liệu thì các thành viên khác của nhóm sẽ không thể đăng ký tài liệu đó được nữa.

Sau khi chú thích xong CTV gửi bài vào hệ thống. Một số bài chú thích sẽ được chấm bởi nhóm quản trị dự án để làm cơ sở đánh giá chất lượng công việc của CTV. Trước khi bài được chấm, CTV được quyền gửi lại chú thích bao nhiêu lần tùy ý.

Người chấm bài không nhất thiết phải đọc bài từ đầu đến cuối và kiểm tra từng chú thích mà chỉ cần sử dụng công cụ so sánh để phát hiện điểm khác nhau giữa hai bản chú thích. Những chỗ có sự mâu thuẫn giữa hai người chú thích có khả năng sai cao hơn nên làm cách này người chấm có thể chấm nhanh hơn tốc độ chú thích nhiều lần mà vẫn đảm bảo độ chính xác chấp nhận được. Chấm xong người chấm bài gửi bài lên hệ thống, có thể gửi lại nhiều lần. Tất cả mọi người đều có thể xem và tải về bài chấm.

Nhóm quản trị có thể tải về toàn bộ chú thích, tải về tài liệu, các bài chú thích được trộn với nhau để chấm hoặc một số kết quả tổng hợp khác để phục vụ việc theo dõi và trích rút thông tin.

Một kinh nghiệm rút ra từ đợt chú thích đầu tiên là nên thường xuyên trích xuất các từ tìm được (vd Feature, Product, Opinion Expression) để xử lý sớm các từ không hợp lý. Mặc dù độ đồng thuận không thấp (60-80% tùy tiêu chí) nhưng kết quả đợt đầu tiên cho khá nhiều từ không chính xác.

### ePi-Teamware: Công cụ cộng tác và quản lý xây dựng ngữ liệu

#### Chức năng

| Chức năng                | Lớp điều khiển               | Ghi chú |
|--------------------------|------------------------------|---------|
| Trang chủ                | HomeController               |         |
| Xác thực                 |                              |         |
| Đăng nhập                | LoginFormController          |         |
| Đăng xuất                | LogoutController             |         |
| Tạo người dùng           | UserCreateFormController     | admin   |
| Sửa thông tin người dùng | UserEditFormController       | admin   |
| Tài liệu                 |                              |         |
| Danh sách                | DocumentListController       |         |
| Xem                      | DocumentViewController       |         |
| Tìm kiếm                 | DocumentSearchFormController |         |
| Xoá                      | DocumentDeleteController     |         |
| Chú thích                |                              |         |
| Danh sách                | AnnotationListController     |         |
| Xem                      | AnnotationViewController     |         |

|                                  |   |   |
|----------------------------------|---|---|
| Tìm kiếm                         | AnnotationSearchFormController              |   |
| Nhận bài để<br>chú thích         |   |   |
| Gửi chú<br>thích                 | AnnotationSubmitFormController              |   |
| Gửi lại chú<br>thích             | AnnotationResubmitFormController            | nếu chưa chấm                                     |
| Kiểm tra chú<br>thích            | AnnotationValidateFormController            | kiểm tra nhưng không gửi                          |
| Kiểm tra<br>toàn bộ chú<br>thích | AnnotationBatchValidateController           | admin   |
| <i>Chấm bài</i>                  |   |   |
| Chấm chú<br>thích                | AnnotationScoreFormController               |   |
| Chấm tài<br>liệu                 | DocumentScoreFormController                 | chấm đồng thời hai chú thích<br>của cùng tài liệu |
| <i>Thống kê</i>                  |   |   |
| Chú thích                        | AnnotationStatisticsController              | theo người chú thích, theo<br>nhóm                |
| Chấm bài                         | ScoringStatsController                      | theo người chú thích, theo<br>người chấm          |
| <i>Tổng hợp<br/>kết quả</i>      |   |   |
| Tất cả chú<br>thích              | AnnotationDownloadAllController             |   |
| Chú thích<br>chưa chấm           | AnnotationDownloadUnscoredController        |   |
| Chú thích<br>trộn để tính<br>IAA | AnnotationDownloadCombiningGroupsController |   |
| Kết quả chú<br>thích             | AnnotationDownloadFinalController           |   |
| <i>Tính lương</i>                |   |   |
| Bảng lương<br>HTML               | UserViewController                          | trang cá nhân của annotator                       |
| Xuất bảng<br>lương CSV           | UserSalaryReportController                  | URL: usersalary.html?id=...                       |
| Trộn thư                         | MailMergeController                         | mailmerge.html                                    |
| Nhật ký                          | LogListController                           | log.html  |

#### *Thiết kế*

- Mô hình: MVC
- Framework: Spring
- Application container: Tomcat
- CSDL: MySQL
- Ảnh xạ đối tượng - quan hệ: Hibernate

Mã nguồn

Truy cập: <http://192.168.1.55:81/hg/epi-teamware/graph>.

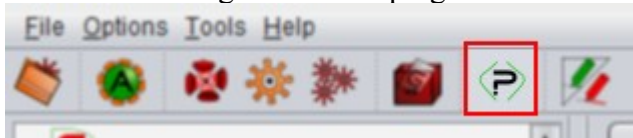
Triển khai (deploy)

- Máy phục vụ: 210.211.108.194
- Tomcat: /root/tomcat - cổng 8080
- Webapp: epi-teamware
- URL: 210.211.108.194:8080/epi-teamware

## ReferenceSchemaAnnotationEditor: Công cụ chú thích

Nạp plugin

1. Nhấn nút “Manage CREOLE plugins” trên thanh công cụ.



2. Trong hộp thoại mở ra, nhấn nút “Add a CREOLE repository”.
3. Chọn thư mục ReferenceEditor > Open > OK và nhấn vào ô “Load now” trên dòng mới được thêm.
4. Lặp lại bước trên với thư mục “schemas”.

Chú thích các thực thể và tính năng

Những cụm từ nói đến cùng một thực thể hợp thành một chuỗi đồng tham chiếu. Người chú thích cần tham chiếu tất cả các thành viên đến phần tử đầu tiên của chuỗi đó.

VD: Trong một tài liệu nhắc đến Asus nhiều lần, lần đầu tiên ta bỏ trống thuộc tính co-reference, các lần tiếp theo ta chọn co-reference đến chú thích đầu tiên.

Trong công cụ chú thích, các thực thể được biểu diễn bằng nút bấm (toggle button) liệt kê theo **thứ tự bảng chữ cái**. Nếu số lượng thực thể quá nhiều thì danh sách thực thể chuyển sang dạng combobox.

Bằng việc tung ra hàng loạt sản phẩm laptop cao cấp, **Asus** không ngần ngại tuyên bố, sẽ trở thành một trong 5 tên tuổi laptop hàng đầu Việt Nam trong vòng 1 năm tới.

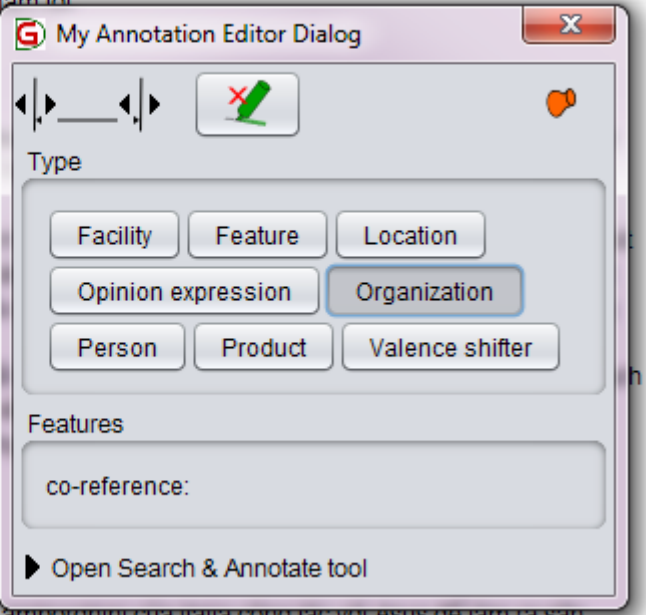
Asus U1F.

Loạt sản phẩm mới như Asus Lamborghini VX2, A trên khắp thế giới, nhưng ở Việt Nam tất cả mới chỉ

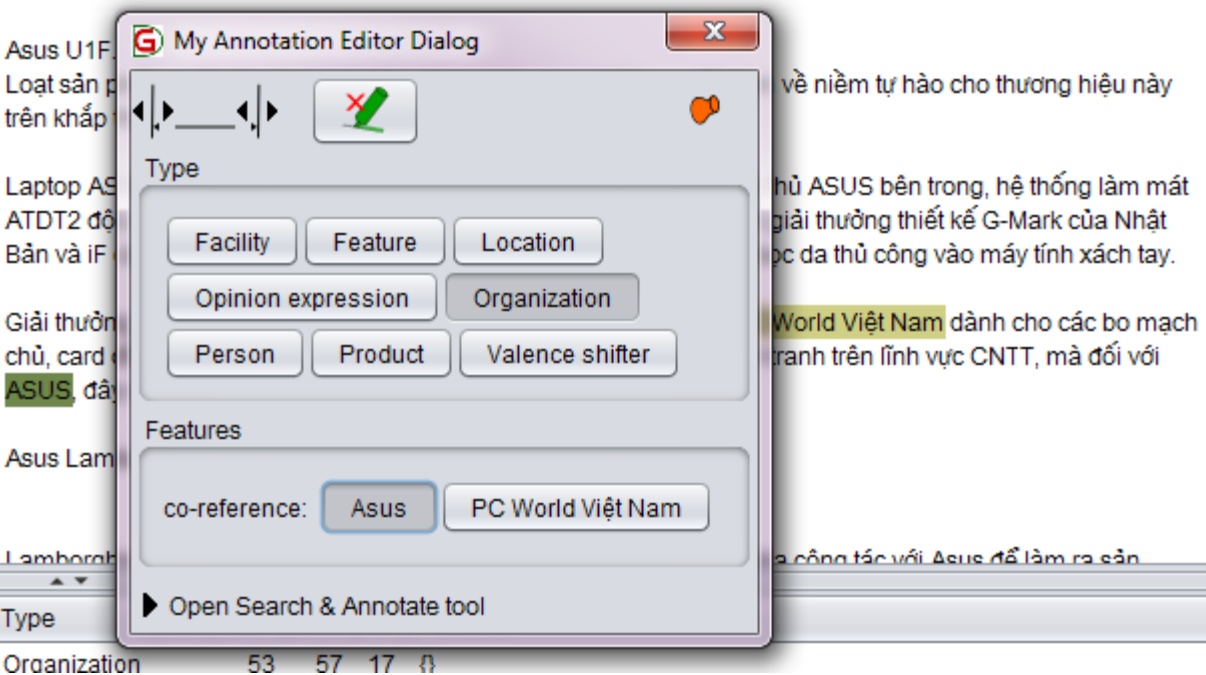
Laptop ASUS được thiết kế mỏng và nhẹ, và ổn định ATDT2 độc quyền, sự cơ động và sự bảo mật an toàn. Bản và iF cùng RedDot của Đức, ASUS là hãng đầu

Giải thưởng "Sản phẩm CNTT được ưa chuộng nhất" chủ, card đồ họa, ổ đĩa quang không phải là đoạn ASUS, đây chỉ mới là sự khởi đầu. Dưới đây là nhữ

Asus Lamborghini VX2



Bằng việc tung ra hàng loạt sản phẩm laptop cao cấp, **Asus** không ngần ngại tuyên bố, sẽ trở thành một trong 5 tên tuổi laptop hàng đầu Việt Nam trong vòng 1 năm tới.



*Chú thích các tính năng, ý kiến và từ chuyển nghĩa*

Mỗi loại này đều chứa tham chiếu đến chú thích khác, trong hầu hết trường hợp người chú thích cần điền tham chiếu này thì chú thích mới có ý nghĩa (chi tiết xem phần Các loại chú thích).

Danh sách các tham chiếu được liệt kê theo nguyên tắc:

- **Không trùng lặp:** trong một dãy đồng tham chiếu chỉ hiện chú thích gần nhất với chú thích

hiện tại.

- **Xa dần:** các chú thích xuất hiện theo thứ tự khoảng cách đến chú thích đang sửa tăng dần.  
**Trong hầu hết trường hợp người chú thích chỉ cần nhìn vào nút đầu tiên** của danh sách vì một cụm từ thường tham chiếu đến đối tượng gần nó nhất.

Hiện tại công cụ không hỗ trợ tham chiếu đến nhiều đối tượng cùng lúc nên nếu có tham chiếu đến một cụm đối tượng thì **chỉ cần chọn một đối tượng trong đó để đại diện**.

VD: M60 (Product) và M90 (Product) kết nối với máy tính qua cổng USB, cắm vào là sử dụng ngay (mình dùng Windows 7), rất nhANH (Opinion expression, opinion-of --> "M90").

Lamborghini VX2 được đội ngũ hãng sản xuất xe Lamborghini của Italia cộng tác với Asus để làm ra sản phẩm này nên máy có rất nhiều chi tiết mang dấu ấn của hãng xe lừng danh này. Chiếc máy tính xách tay này được thiết kế giống hệt phần khung hậu của chiếc xe đua Lamborghini Miura 1970, màu sơn và chất liệu sơn giống hệt sơn xe.

Chiếc máy có cấu hình mạnh: bộ vi xử lý Intel Core 2 Duo T7500 2x2.2G Santa Rosa; ram 2Gb, ổ đĩa cứng 200Gb, card đồ họa NVIDIA GeForce 8400 GS 128Mb. Sản phẩm này được phân phối vào loại "siêu

Asus S6Fm

Model Asus S6Fm  
Asus, vì làm việc  
quyết tốt vấn đề  
này nên trông

Chiếc máy hươ  
chuẩn, nó chỉ

Type

Organization  
Organization  
Organization  
Product  
Product  
Feature

My Annotation Editor Dialog

Navigation: Previous, Next, Home, Search, Cancel, Pin

Type

Facility Feature Location Opinion expression  
Organization Person Product Valence shifter

Features

co-reference: Chiềc máy ASUS PC World Việt Nam

feature-of: RedDot

Open Search & Annotate tool

vào loại "siêu

lý kỹ. Theo  
Đa Asus đã giải  
ng tay lớp da

và pin 6cell tiêu  
M.1.33G ram

1303 1311 25 {feature-of=24}

Chiếc máy có cấu hình mạnh: bộ vi xử lý Intel Core 2 Duo T7500 2x2.2G Santa Rosa; ram 2Gb, ổ đĩa cứng 200Gb, card đồ họa nVidia GeForce 512Mb, tích hợp các công nghệ tân kỳ nhất. Giá bán cũng vào loại "siêu khủng": 3.388 USD

**My Annotation Editor Dialog**

Navigation icons: Previous, Next, Home, End, Cancel (X), Save (Floppy), Pin.

Type

Facility Feature Location Opinion expression  
Organization Person Product Valence shifter

Features

polarity\*: NEGATIVE NEURAL POSITIVE

feature:

opinion-of: Giá bán cấu hình Chiếc máy  
ASUS PC World Việt Nam RedDot

► Open Search & Annotate tool

đã được xử lý kỹ. Theo các nhà thiết kế của Asus đã giải được gia công bằng tay lớp da

hình 11,1inch và pin 6cell tiêu 2 Duo L7200 LV 1.33G ram

TIVE}

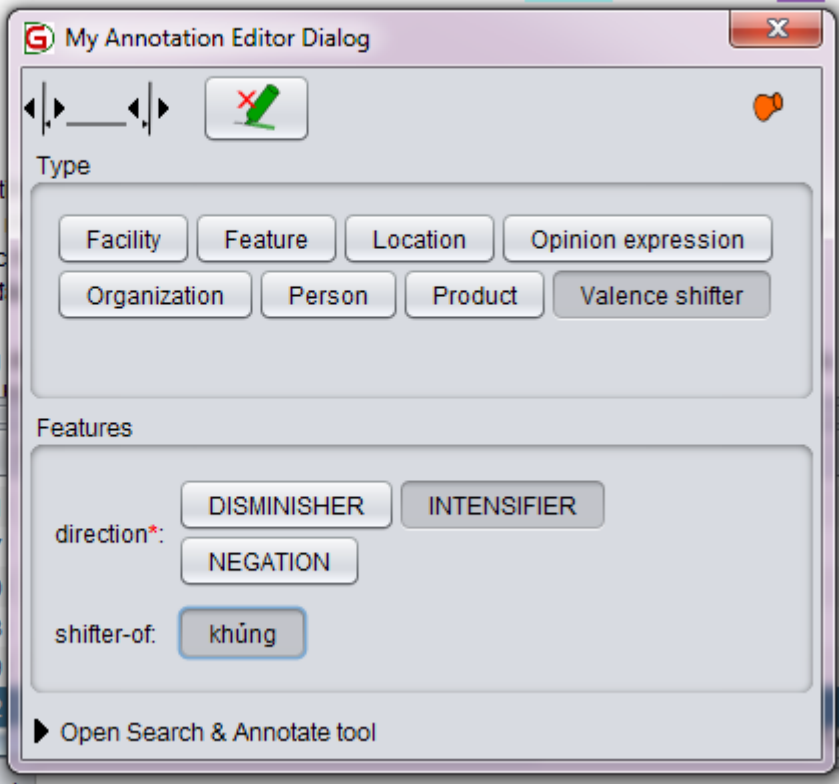
Chiếc máy có cấu hình mạnh: bộ vi xử lý Intel Core 2 Duo T7500 2x2.2G Santa Rosa; ram 2Gb, ổ đĩa cứng 200Gb, card đồ họa nVidia GeForce 512Mb, tích hợp các công nghệ tân kỳ nhất. Giá bán cũng vào loại "siêu khủng": 3.388 USD

Asus S6Fm

Model Asus S6Fm là dạng máy t Asus, vì làm vỏ bằng da nên tản quyết tốt vấn đề tản nhiệt và thực này nên trông rất tinh tế và độc đ

Chiếc máy hướng đến đối tượng chuẩn, nó chỉ nặng 1.53kg. Cấu

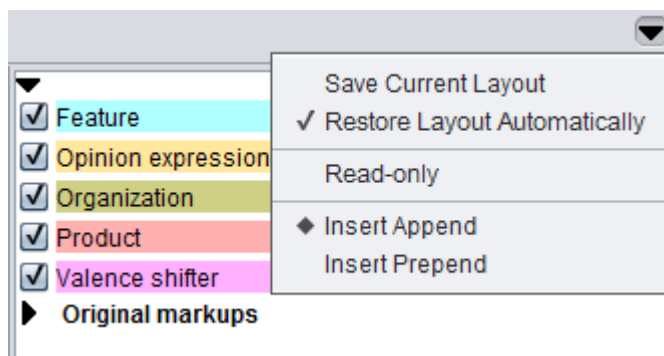
| Type            | Set | Start |
|-----------------|-----|-------|
| Organization    |     | 861   |
| Product         |     | 957   |
| Product         |     | 1290  |
| Feature         |     | 1303  |
| Feature         |     | 1469  |
| Valence shifter |     | 1492  |



Kinh nghiệm chủ thích hiệu quả

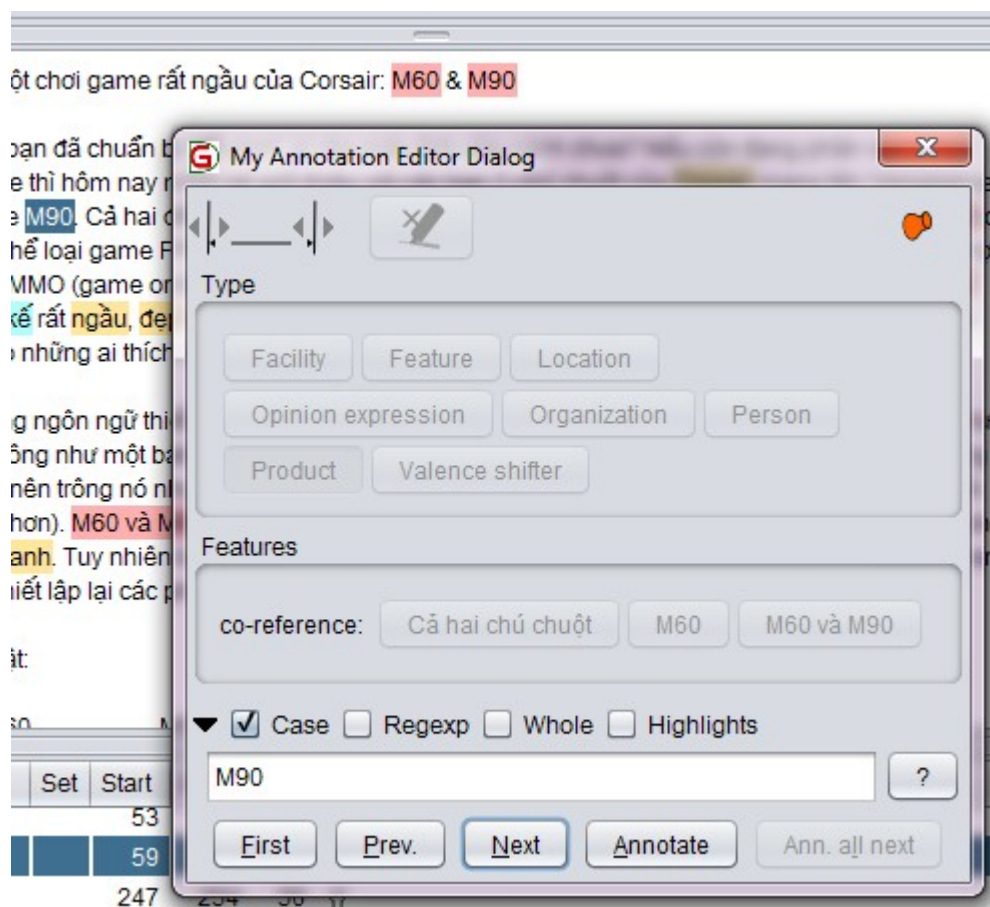
1. Bật **chế độ chỉ đọc** để tránh thay đổi nội dung bài viết (mục "Read-only" trong hình dưới)
2. Thứ tự chú thích: **Product > Feature > Opinion expression > Valence shifter**.
3. Đóng bộ soạn thảo nhanh: nhấn phím **Esc**.

4. Tự động hiển thị các loại chú cần dùng khi mở tài liệu mới:
  - a. Bật các loại chú thích muốn hiển thị.
  - b. Nhấn nút mũi tên bên phải khung Annotation Sets tay phải màn hình làm để mở thực đơn ngữ cảnh.
  - c. Bật chức năng **"Restore layout automatically"** nó chưa được bật.



thích  
mong  
(bên  
việc)  
nếu

- d. Mở thực đơn vừa nảy một lần nữa và nhấn **"Save current layout"**.
5. Tìm và chú thích: Chức năng tiện lợi khi chú thích tên riêng.
    - a. Đánh dấu lần xuất hiện đầu tiên của một tên riêng.
    - b. Nhấn vào dòng chữ **"Open search & annotate tool"**
    - c. Nhấn nút Next để tìm tên đó trong tài liệu, kiểm tra để đảm bảo không chú thích sai, sót.
    - d. Nhấn Annotate để đánh dấu.



### Công cụ so sánh

Khi chú thích tài liệu có tham chiếu, một khó khăn đặt ra là nhìn vào mã hiệu của tham chiếu ta không biết nó có trỏ đến đúng đối tượng hay không. Công cụ Reference Annotation Differ (trong dự án vnlp, thư mục gate/Reference\_Utils) thay thế công cụ so sánh mặc định của GATE để hiển thị thêm thông tin này trên hộp thoại đối sánh giúp người chấm bài làm việc hiệu quả hơn. Tương lai cần thêm chức năng so sánh trường tham chiếu để chấm bài chính xác hơn.