ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

BÁO CÁO KỸ THUẬT

HỆ THỐNG TÓM TẮT SỰ KIỆN

Người thực hiện: Trần Mai Vũ

Đơn vị công tác: Phòng Thí nghiệm Công nghệ Tri thức,
Trường Đại học Công nghệ,
Đại học Quốc gia Hà Nội

Hà Nội, 8-2019

Mục lục

1. Giới thiệu về tóm tắt sự kiện

Tóm tắt văn bản là một vấn đề nhận được nhiều sự quan tâm của cộng đồng nghiên cứu với rất nhiều các bài báo khoa học xuất hiện tại các hội nghị lớn như: ACL, DUC, TAC, COLING, EMNLP... Mục đích chính của tóm tắt văn bản đó chính là đưa ra văn bản ngắn gọn và xúc tích hơn từ một văn bản đầu vào. Bài toán này cũng được ứng dụng nhiều trong các hệ thống thực tế như: tóm tắt các trang web trong máy tìm kiếm, tóm tắt các tin tức, tóm tắt các quan điểm người dùng… Một trong những bài toán đang được quan tâm hiện nay đối với tóm tắt đó chính là **tóm tắt sự kiện**.

Tóm tắt sự kiện là một bài toán với đầu vào là một cụm các sự kiện có liên quan đến nhau và đầu ra là một văn bản rút gọn mô tả về cụm sự kiện đấy. Có khá nhiều các công trình nghiên cứu về tóm tắt sự kiện áp dụng trên nhiều miền dữ liệu khác nhau như: tóm tắt sự kiện trên mạng xác hội Twitter [NSS09, CP11, SKW13, CA13] hay tóm tắt các sự kiện tin tức [WLL07, WWL10, LLW07, Gi13]. Trong báo cáo này chúng tôi tập trung vào vấn đề "tóm tắt các sự kiện tin tức liên quan đến thực thể tên người trên miền dữ liệu tiếng Việt".

Bài toán tóm tắt sự kiện là một bài toán khó do kết quả phụ thuộc vào khá nhiều các bài toán khác trong lĩnh vực xử lý ngôn ngữ tự nhiên như: nhận dạng thực thể, trích chọn quan hệ, phát hiện đồng tham chiếu, phân tích cây cú pháp… Có nhiều hướng tiếp cấn để xây dựng mô hình tóm tắt sự kiện, tuy nhiên các phương pháp thường chia bài toán này thành hai phần chính là trích chọn sự kiện và tóm tắt sự kiện.

2. Sự kiện và trích chọn sự kiện

2.1. Các định nghĩa về sự kiện

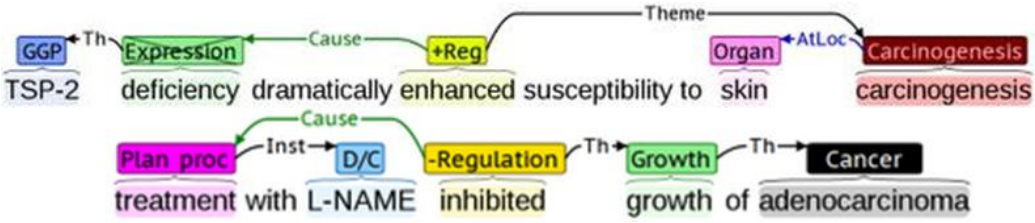
Tùy theo từng lĩnh vực và dữ liệu người ta có nhiều cách định nghĩa sự kiện:

- Trên miền tin tức, Allan và cộng sự định nghĩa tin tức chứa sự kiện nếu nó có bốn yếu tố: hành vi, chủ thể, thời gian và địa điểm [APL98].
- Hội nghị MUC quan tâm đến các sự kiện về khủng bố, quân sự, đầu tư mạo hiểm, tai nạn máy bay… Định nghĩa sự kiện mà hội nghị đưa ra phải có đủ các yếu tố: tác nhân, thời gian, địa điểm và các tác động của nó.
- Chương trình ACE[1] (Automatic Content Extraction), sự kiện đơn giản là một sự thay đổi trạng thái. Loại sự kiện và các thuộc tính sự kiện được quy định chặt chẽ hơn. Có tám loại sự kiện được sử dụng bao gồm business (kinh tế), conflict (xung đột), contact (liên lạc), justice (pháp lý), life (cuộc sống), movement (sự di chuyển), personnel (nhân sự) và transaction (giao dịch). Mỗi loại sự kiện sau đó lại được chia thành từng dạng con. Ví dụ như trong justice bao gồm một số dạng như arrest – jail (bắt giữ – bỏ tù), convict (kết án), fine (phạt)…[Ah06]
- Trong hệ thống VnLoc [VHQ12] sự kiện được định nghĩa là bộ bảy đặc trưng bao gồm tên sự kiên, loại sự kiện, thời gian xảy ra sự kiện, nơi xảy ra sự kiện, nguồn đưa tin, liên kết và tóm tắt của sự kiện đó. Cũng theo VnLoc thì sự kiện họ quan tâm thuộc một trong ba loại: tai nạn giao thông, hình sự, cháy nổ.

[1] http://projects.ldc.upenn.edu/ace/

Hình 1. Minh họa các sự kiện có trong hệ thống VNLOC

- Trong trích xuất các sự kiện y sinh, sự kiện y sinh là một thay đổi về trạng thái của một hoặc nhiều phân tử sinh học. Nó được định nghĩa như một quan hệ giữa một hoặc nhiều thực thể thực hiện các vai trò khác nhau [CHR05, CVJ09].



Hình 2. Một ví dụ các sự kiện y sinh xuất hiện trong một câu

2.2. Định nghĩa sự kiện trong các bài báo tin tức tiếng Việt

Sau khi khảo sát các định nghĩa trong các miền dữ liệu khác nhau của tiếng Anh, chúng tôi nhận thấy:

- Định nghĩa của Allan cho chúng ta cái nhìn tổng quát về sự kiện trong miền dữ liệu tin tức với bốn yếu tố chính là: hành vi, chủ thể, thời gian và địa điểm.

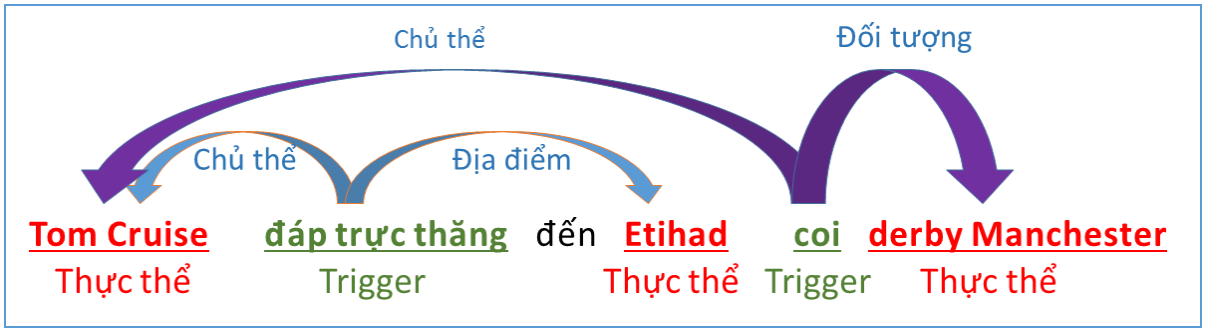
- Định nghĩa sự kiện trong miền dữ liệu y sinh thể hiện rõ các mối quan hệ ngữ nghĩa giữa các thành phần của sự kiện.

Từ hai định nghĩa trên, chúng tôi đưa ra một định nghĩa về sự kiện tin tức trong tiếng Việt: Sự kiện tin tức trong tiếng Việt mô tả một hành vi cụ thể của một chủ thể tại một thời gian và địa điểm xác định. Hành vi này được biểu diễn bằng một mối quan hệ ngữ nghĩa giữa một hay nhiều thực thể thực hiện cùng một vai trò với nhau.

Từ định nghĩa sự kiện y sinh học ta có các khái niệm liên quan:

- **Thực thể**: là danh từ hoặc cụm danh từ biểu diễn cho chủ thể và các tác nhân liên quan. Ví dụ như thực thể tên người, thực thể tên tổ chức
- **Trigger sự kiện**: từ hoặc cụm từ trong câu, chỉ ra sự xuất hiện của sự kiện và mang kiểu của sự kiện. Ví dụ: trong sự kiện "Tai nạn giao thông" ta có thể có các trigger sự kiện như: tông xe, đâm xe, lao xuống vực,...
- **Tham số** (của sự kiện): các thực thể hoặc các sự kiện khác góp phần mô tả sự kiện, cũng là một phần của sự biểu diễn sự kiện, và thường được phân loại theo vai trò ngữ nghĩa.
- **Vai trò tham số:** mỗi một tham số biểu diễn một sự tác động hoặc tham gia vào sự kiện của một thực thể.

Ví dụ dưới đây thể hiện rõ sự xuất hiện của sự kiện và thành phần trong câu:



Hình 3. Ví dụ về sự xuất hiện của sự kiện và thành phần trong câu

Trong câu trên chúng ta thấy có hai sự kiện xuất hiện là "Tom Cruise đáp trực thăng đến Etihad" và "Tom Cruise coi derby Manchester". Các cụm từ bôi đỏ

là các thực thể và các cụm từ bôi xanh là các Trigger. Mỗi một sự kiện có hai quan hệ liên quan đến Trigger mỗi một quan hệ thể hiện mối quan hệ giữa trigger và thực thể, ở đây thực thể là tham số của sự kiện và nhãn của quan hệ sẽ là vai trò của tham số.

Chúng tôi đưa thêm hai khái niệm liên quan đến sự kiện là:

- **Bị động:** chỉ các sự kiện mà chủ thể đóng vai trò bị động ở trong câu, ví dụ: "Quốc Trung bị chê"
- **Phủ định:** chỉ sự phủ định đối với một sự kiện, ví dụ: "Đàm Vĩnh Hưng không xuất hiện tại Hà Nội" ở đây sự kiện "xuất hiện" được đánh dấu là phủ định.

2.3. Trích chọn sự kiện

Trích chọn sự kiện được cộng đồng khoa học quốc tế đầu tư nghiên cứu từ khá sớm. Hội nghị MUC[2] được tổ chức lần đầu tiên năm 1987 dưới sự hỗ trợ của Quỹ nghiên cứu bộ quốc phòng Hoa Kỳ là một trong những hội nghị tiêu biểu trong trích chọn sự kiện. Hội nghị đã đưa ra phương pháp trích chọn sự kiện theo khung mẫu với mục đích là trích chọn bằng cách lấy các thông tin liên quan đến sự kiện. Bên cạnh đó, các chương trình TDT[3] (Phát hiện và theo dõi chủ đề) được tổ chức hàng năm từ năm 1997 đã bước đầu giải quyết được bài toán phát hiện sự kiện mới, theo dõi và xâu chuỗi sự kiện. Có nhiều nhóm nghiên cứu tham gia chương trình như nhóm BBN từ công ty BBN Technologies, nhóm CMU của trường đại học Carnegie Mellon, nhóm DRAGON của công ty Dragon Systems… Mỗi nhóm đều đưa ra những tiếp cận riêng và góp phần nâng cao kết quả của lĩnh vực trích chọn sự kiện.

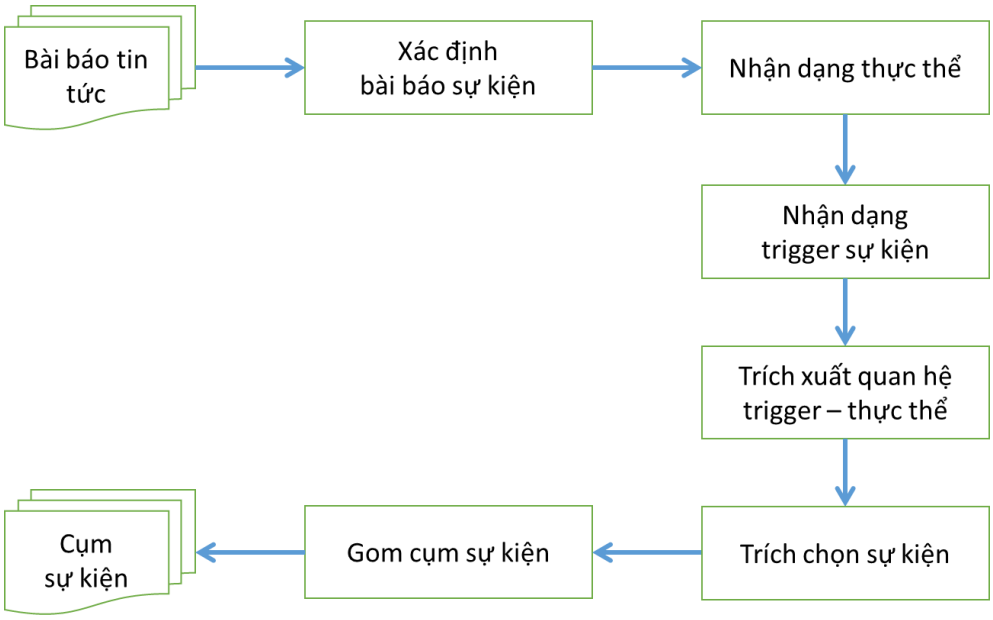
Có hai mô hình thường được sử dụng trong các phương pháp trích xuất sự kiện đó là mô hình pipeline và mô hình suy luận gộp (joint inference), cả hai mô hình đều có những ưu nhược điểm riêng. Các mô hình suy luận gộp sử dụng một số phương pháp như biểu

[2] http://www-nlpir.nist.gov/related_projects/muc/

[3] http://projects.ldc.upenn.edu/TDT/

diễn sự kiện dưới dạng cây cú pháp rồi phân tích hay sử dụng phương pháp suy luận Markov logic, các mô hình này cho hiểu qua tương đối tốt tuy nhiên phức tạp trong việc biểu diễn và xử lý dữ liệu. Mô hình pipeline là mô hình tuần tự chia các phần của bài toán trích chọn sự kiện thành các pha nhỏ để giải quyết, phương pháp này giúp cho việc quản lý và phát triển mô hình được dễ dàng hơn. Chúng tôi quyết định sử dụng mô hình pipeline cho mô hình trích chọn sự kiện tin tức tiếng Việt.

Dưới đây là mô hình trích chọn sự kiện tin tức tiếng Việt:



Hình 4. Mô hình trích chọn sự kiện tin tức

- **Đầu vào:** Các bài báo tin tức tiếng Việt
- **Đầu ra:** Các cụm sự kiện liên quan đến chủ thể là người

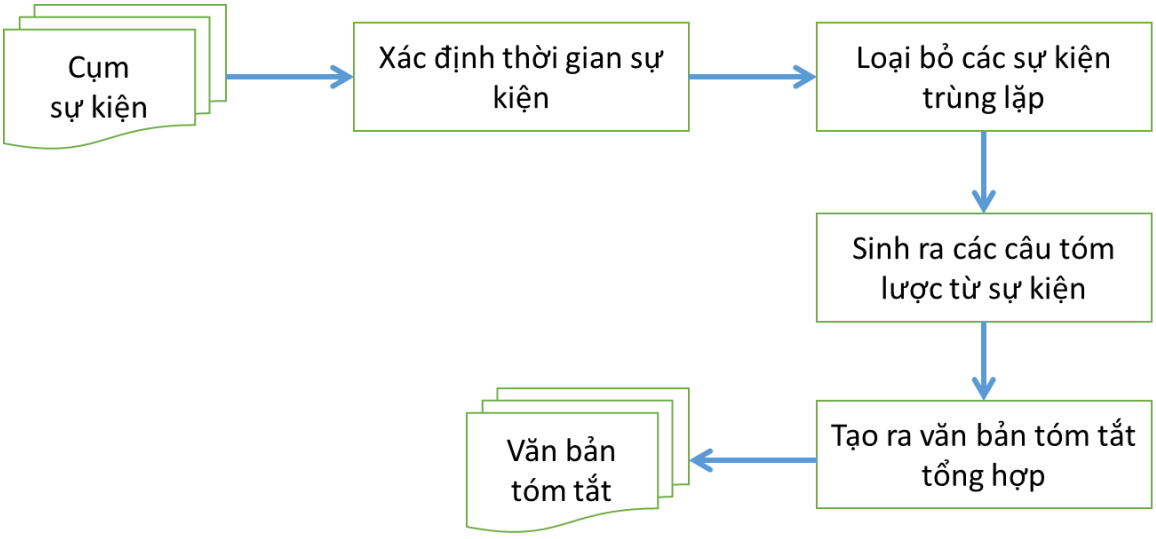
Các pha của mô hình:

- **Xác định sự kiện**: Các bài báo được đưa vào mô hình sẽ được đưa qua một bộ phân lớp nhị phân nhằm xác định xem bài báo đấy có phải viết về một sự kiện hay không. Chúng tôi chỉ sử dụng phần tiêu đề và tóm tắt của bài báo để tiến hành phân lớp.

- **Nhân dạng thực thể**: Trong pha xử lý này các thực thể liên quan đến sự kiện sẽ được trích xuất, chúng tôi áp dụng phương pháp học máy Maximum Entropy giải mã bằng thuật toán Beam search kết hợp với nhận dạng bằng từ điển dựa trên phương pháp longest matching để nhận dang ra các thực thể.
- **Nhận dạng trigger sự kiện**: Các trigger được nhận dạng dựa trên một tập từ vựng các trigger đã được nhóm xây dựng sẵn.
- **Trích xuất quan hệ trigger – thực thể**: tiến hành ghép cặp các trigger và thực thể đã nhận dạng được, các cặp này sẽ được đưa qua một bộ phân lớp nhị phân để xác định xem cặp nào sẽ được giữ lại và cặp nào sẽ bị loại bỏ. Mô hình phân lớp được học trên một tập dữ liệu đã được gắn nhãn thủ công.
- **Trích chọn sự kiện**: các cặp trigger - thực thể được giữ lại trong câu và liên quan trực tiếp đến một chủ thể sẽ được tổng hợp thành sự kiện.
- **Gom cụm sự kiện**: các sự kiện liên quan đến nhau cần được gom nhóm lại thành một cụm. Các cụm liên quan đến nhau thường được chỉ về cùng một chủ thể và có một sự tương đồng về các đối tượng liên quan, bên cạnh đấy các sự kiện này thường xảy ra trong cùng một thời gian nhất định (ví dụ trong một ngày). Chúng tôi tiến hành xây dựng một mô hình gom cụm các sự kiện dựa vào một tập luật để lọc các sự kiện gần nhau và phương pháp gom cụm Nearest Neighbor Clustering.

3. Tóm tắt sự kiện

Trong phần tóm tắt sự kiện, chúng tôi tiến hành sinh các văn bản tóm tắt cho các cụm sự kiện đầu ra. Mô hình tóm tắt bao gồm các bước sau:



Hình 5. Mô hình tóm tắt sự kiện

- **Đầu vào:** Cụm sự kiện đầu ra của trích xuất sự kiện
- **Đầu ra:** Văn bản tóm tắt cho cụm sự kiện đấy

Các pha của mô hình:

- **Xác định thời gian sự kiện:** Sử dụng một số luật dựa trên biểu thức chính quy và từ điển, chúng tôi nhận diện các cụm thời gian có trong câu có liên quan đến sự kiện. Ví dụ: "Hôm nay, Kiến Huy lần đầu song ca cùng Như Quỳnh", cụm từ chỉ thời gian "hôm nay" có quan hệ với sự kiện "lần đầu song ca".
- **Loại bỏ các sự kiện trùng lặp:** Trong cụm sự kiện, các sự kiện có cùng chủ thể, trigger sự kiện và xuất hiện trong cùng một khoảng thời gian sẽ được xem như sự kiện trùng lặp, các sự kiện trùng lặp sẽ được loại bỏ. Việc xác định trùng lặp tương đối phức tạp do xuất hiện các từ/cụm từ đồng

nghĩa trong trigger sự kiện hoặc thực thể, chúng tôi sử dụng một từ điển đồng nghĩa để xử lý vấn đề này tuy nhiên số lượng giải quyết được vẫn rất hạn chế.

- **Sinh ra các câu tóm lược từ sự kiện:** Các sự kiện thường xuất hiện trong tin tức tiếng Việt thường nằm ở dạng có một hoặc hai tham số. Ví dụ:
    - Sự kiện một tham số trong câu "Thái Trinh giảm 7 kg để hát nhạc Dance"

      < **Thái Trinh** (Tham số vai trò chủ thể) ← **giảm 7 kg** (Trigger sự kiện)>
    - Sự kiện hai tham số tham số trong câu "Hương Tràm hóa cáo để thân thiết với Cao Thái Sơn"

< **Hương Tràm** (Tham số vai trò chủ thể) ← **thân thiết** (Trigger sự kiện) → **Cao Thái Sơn** (Tham số vai trò đối tượng)>

Chúng tôi sinh ra các câu ngắn gọn hơn bằng cách kết hợp các tham số, trigger sự kiện và các giới từ nằm giữa chúng, ví dụ:

- "Thái Trinh giảm 7 kg để hát nhạc Dance" → "Thái Trinh giảm 7 kg"
- "Hương Tràm hóa cáo để thân thiết với Cao Thái Sơn" → "Hương Tràm thân thiết với Cao Thái Sơn"

- **Tạo văn bản tóm tắt tổng hợp:** Từ các câu ngắn gọn và thời gian của từng sự kiện, hệ thống tạo ra văn bản tóm tắt tổng hợp cho từng chủ thể có trong hệ thống, việc tạo ra văn bản tóm tắt dựa trên việc sắp xếp các câu tóm lược theo trình tự thời gian.

4. Một số kết quả bước đầu

4.1. Xây dựng dữ liệu

Các bước xây dựng dữ liệu:

- Miền dữ liệu được áp dụng là các tin tức thuộc lĩnh vực Văn hóa và Giải trí.
- Thu thập các tin tức thuộc hai lĩnh vực trên từ trang web Baomoi.com.
- Trích xuất các câu từ hai phần của tin tức là tiêu đề và tóm tắt.
- Dựa trên nhưng câu đã thu được, tiến hành gắn nhãn bằng tay các sự kiện và thành phần sự kiện xuất hiện trong các câu
- Xây dựng thủ công các câu tóm lược từ các thành phần sự kiện

	Trigger	Thực thể chính	Thực thể liên quan	Bị động ?	Phủ định ?	Ngữ tóm lược	Câu trọn ven	Liên kết
2	cướp vai	Thành Long	đàn em	bị		Thành Long bị đàn em cướp vai	Thực hư chuyện Thành Long bị đàn em cướp vai	http://www.baomoi.com/Home/SanKhau/ti
3	ngậm trái đắng	Khánh Thy	đàn ông	khiến		Đàn ông khiến Khánh Thy ngậm t	Hai người đàn ông khiến Khánh Thy ngậm trái đắn	http://www.baomoi.com/Home/AmNhac/pl
4	cưỡng ôm	Huỳnh Thanh Y		bị		Huỳnh Thánh Y bị cưỡng ôm	Huỳnh Thánh Y bị cưỡng ôm	http://www.baomoi.com/Home/SanKhau/gi
5	chia tay	Angelababy	Huỳnh Hiểu Minh			Angelababy chia tay Huỳnh Hiểu	Angelababy "làm lơ" trước tin chia tay Huỳnh Hiể	http://www.baomoi.com/Home/SanKhau/at
6	giảm 7 kg	Thái Trinh				Thái Trinh giảm 7 kg	Thái Trinh giảm 7 kg để hát nhạc Dance	http://www.baomoi.com/Home/AmNhac/ke
7	phấn khích	Đàm Vĩnh Hưng	clip			Đàm Vĩnh Hưng phấn khích với cl	Đàm Vĩnh Hưng phấn khích với clip cậu bé 10X cov	http://www.baomoi.com/Home/AmNhac/sc
8	đi chơi	Suri Cruise	mẹ			Suri Cruise đi chơi với mẹ	Suri Cruise vui vẻ nhảy chân sáo đi chơi cùng mẹ	http://www.baomoi.com/Home/SanKhau/at
9	nổi loạn	Đỗ Hoàng Diệu				Đỗ Hoàng Diệu nổi loạn	Đỗ Hoàng Diệu có còn nổi loạn?	http://www.baomoi.com/Home/SachBaoVa
10	song ca	Kiến Huy	Như Quỳnh			Kiến Huy song ca Như Quỳnh	Kiến Huy lần đầu song ca cùng Như Quỳnh	http://www.baomoi.com/Home/AmNhac/w
11	hát hay	Johnny Trí Nguyễn				Johnny Trí Nguyễn hát hay	Điên đảo vì Johnny Trí Nguyễn hát hay	http://www.baomoi.com/Home/AmNhac/w
12	thân thiết	Hương Tràm	Cao Thái Sơn			Hương Tràm thân thiết Cao Thái	Hương Tràm hóa cáo để thân thiết với Cao Thái Sc	http://www.baomoi.com/Home/AmNhac/pl
13	tham gia	Adam Lambert	Glee			Adam Lambert tham gia Glee	Adam Lambert tham gia Glee mùa thứ 5	http://www.baomoi.com/Home/AmNhac/vt
14	bắt tại trận	Kim Tử Long	chiếu bạc	bị		Kim Tử Long bị bắt tại trận tại chi	NSƯT Kim Tử Long bị "bắt tại trận" trên một chiếu	http://www.baomoi.com/Home/AmNhac/gi
15	vướng nghi án	Bùi Anh Tuấn	ma túy đá	bị		Bùi Anh Tuấn bị vướng nghi án m	ca sĩ trẻ Bùi Anh Tuấn bị vướng nghi án sử dụng m	http://www.baomoi.com/Home/AmNhac/gi
16	thừa hưởng nét	Angelina Jolie	mẹ	được		Angelina Jolie được thừa hưởng	Angelina Jolie đã được thừa hưởng tất cả nét đẹp	http://www.baomoi.com/Home/SanKhau/ke
17	nhận lời tham g	Yeon Woo Jin	dự án			Yeon Woo Jin nhận lời tham gia	Chàng nam thứ vừa gây sóng gió trong cộng đồng	http://www.baomoi.com/Home/SanKhau/ke
18	tham gia	Yeon Woo Jin	dự án			Yeon Woo Jin tham gia dự án	Chàng nam thứ vừa gây sóng gió trong cộng đồng	http://www.baomoi.com/Home/SanKhau/ke
19	đưa	Thanh Bùi	chứng chỉ âm nhạc quốc tế			Thanh Bùi đưa chứng chỉ âm nhạ	Thanh Bùi đưa chứng chỉ âm nhạc quốc tế về Việt	http://www.baomoi.com/Home/AmNhac/w

Bảng 1. Ví dụ về các câu tóm lược từ các thành phần sự kiện được xây dựng thủ công

Thống kê dữ liệu đã gắn nhãn:

Số lượng câu	5696
Số lượng sự kiện	7103
Tỷ lệ sự kiện trên câu	1.25
Số lượng trigger sự kiện	3487
Số lượng thực thể chủ thể	1494

Bảng 2. Thống kê dữ liệu đã gán nhãn

4.2. Kết quả phát hiện sự kiện

Xây dựng bộ phân lớp:

- Các câu trong tập đã gắn nhãn được xem như các ví dụ dương (nhãn "Sự kiện"), ghép thêm 6000 câu không phải sự kiện thuộc cùng lĩnh vực Văn hóa – Giải trí (nhãn "Không phải sự kiện").
- Sử dụng phương pháp phân lớp Maximum Entropy với các đặc trưng:
  - Đặc trưng hình thái từ: số, viết hóa, thực thể,…
  - Đặc trưng n-gram các từ
  - Đặc trưng nhãn từ loại
  - Đặc trưng n-gram các nhãn từ loại
  - Đặc trưng từ điển các trigger
- Đánh giá mô hình dựa trên kiểm thử chéo 10 folds (tỷ lệ giữa hai lớp được giữ nguyên 1:1) và độ đo micro F trên nhãn Sự kiện

Nhãn	Độ chính xác	Độ hồi tưởng	Độ đo F
Sự kiện	94.27	86.74	90.35

Bảng 3. Đánh giá mô hình phát hiện sự kiện

4.3. Kết quả trích chọn sự kiện

Các bước thực hiện:

- Sử dụng tập các sự kiện đã được gán nhãn gồm 7103 sự kiện.
- Nhận dạng thực thể sử dụng từ điển và mô hình nhận dạng thực thể tên người được huấn luyện bằng phương pháp Maximum Entropy + Beam search với tập dữ liệu 10.000 câu đã được gắn nhãn thực thể thủ công.
- Nhận dạng các trigger sử dụng tập từ điển trigger và phương pháp Longest matching.
- Việc tổng hợp các thành phần thành sự kiện sử dụng một số luật về khoảng cách và ví trí trong câu.

- Xây dựng bộ phân lớp để loại bỏ các sự kiện dư thừa.
    - Tập huấn luyện được xây dựng với hai lớp, 12.000 câu có được trong thực nghiệm 3.1 sẽ được đưa qua bước trích xuất sự kiện để trích ra các ứng viên sự kiện, 7103 sự kiện đã được gán nhãn sẽ nhận nhãn dương và các sự kiện khác sẽ nhận nhãn âm (khoảng gần 3000 sự kiện mang nhãn âm)
    - Sử dụng phương pháp Maximum Entropy với các đặc trưng:
        - Đặc trưng từ vựng của các thành phần
        - Đặc trưng nhãn từ vựng của các từ trong trigger sự kiện
        - Đặc trưng các từ xung quanh tham số
        - Đặc trưng các từ xung quanh trigger
        - Đặc trưng các từ nằm giữa trigger và các tham số
    - Đánh giá mô hình dựa trên kiểm thử chéo 10 folds (tỷ lệ giữa hai lớp được giữ nguyên 7:3) và độ đo micro F trên nhãn dưỡng (các sự kiện nhãn dương sẽ được giữ lại)

Nhãn	Độ chính xác	Độ hồi tưởng	Độ đo F
Nhãn dương	96.41	92.55	94.44

Bảng 4. Đánh giá mô hình trích chọn sự kiện

5. Công việc tiếp theo

- Thực hiện bước tóm tắt và tổng hợp tóm tắt.
- Xây dựng ứng dụng tóm tắt sự kiện trực tuyến.
- Đánh giá và nâng cấp độ chính xác cho các bước của mô hình

Tài liệu tham khảo

[Ah06] David Ahn. “The stages of event extraction”, In Proceedings of the Workshop on Annotating and Reasoning about Time and Events, 2006, pp. 1-8.

[APL98] James Allan, Ron Papka, and Victor Lavrenko. “On-line new event detection and tracking”, In Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, 1998, pp. 37–45.

[CA13] Freddy Chong Tat Chua, Sitaram Asur. Automatic Summarization of Events from Social Media. 7th International AAAI conference on Weblogs and Social Media Association for the Advancement of Artificial Intelligence (AAAI), 2013

[CHR05] Hong-woo Chun, Young-sook Hwang, and Hae-Chang Rim. “Unsupervised event extraction from biomedical literature using co-occurrence informationand basic patterns”, In Proceedings of the First international joint conference on Natural Language Processing, 2005, pp. 777–786.

[CP11] Deepayan Chakrabarti and Kunal Punera. Event Summarization using Tweets. Association for the Advancement of Artificial Intelligence (AAAI), 2011

[CVJ09] K. Bretonnel Cohen, Karin Verspoor, Helen L. Johnson, Chris Roeder, Philip V. Ogren, William A. Baumgartner, Jr., Elizabeth White, Hannah Tipney, and Lawrence Hunter. “High-precision biological event extraction with a concept recognizer”, In Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task, 2009, pp. 50–58.

[Gi13] Giang Binh Tran. 2013. Structured Summarization for News Events. In 22nd World Wide Web (WWW), Brazil, May 2013

[LLW07] Maofu Liu, Wenjie Li, Mingli Wu and Qin Lu. Extractive Summarization Based on Event Term Clustering. ACL 2007

[NSS09] Yoko Nishihara, Keita Sato, and Wataru Sunayama. “Event extraction and visualization for obtaining personal experiences from blogs”, In Proceedings ofthe

Symposium on Human Interface 2009 on Human Interface and the Management of Information. Information and Interaction. Part II: Held as part of HCI International 2009, pp. 315–324.

[SKW13] Chao Shen, Fei Liu, Fuliang Weng, Tao Li. A Participant-based Approach for Event Summarization Using Twitter Streams, NAACL 2013

[VHQ12] Mai-Vu Tran, Minh-Hoang Nguyen, Sy-Quan Nguyen, Minh-Tien Nguyen, Xuan-Hieu Phan (2012). "VnLoc: A Real–time News Event Extraction Framework for Vietnamese", KSE 2012:161-166, Da Nang, August 17-19, 2012

[WLL07] Mingli Wu, Wenjie Li, Qin Lu, Kam-Fai Wong. Event-Based Summarization Using Time Features. CICLING 2007

[WWL10] Peng Wang, Haixun Wang, Majin Liu, and Wei Wang. 2010. An algorithmic approach to event summarization. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (SIGMOD '10). ACM, New York, NY, USA