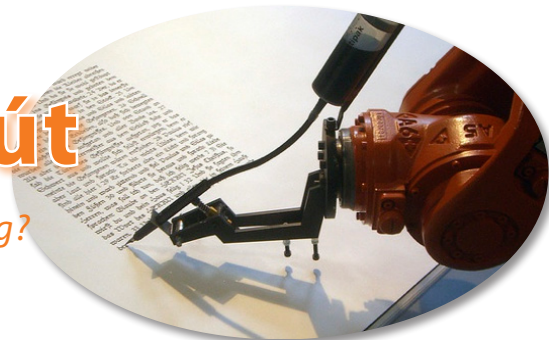


# Khi robot chấp bút

*Nhà báo bị cạnh tranh hay được giải phóng?*



✧ MAI ANH

Một bản tin hoàn tất với tốc độ kỷ lục đã giúp Los Angeles Times trở thành tờ báo đầu tiên đưa tin về trận động đất tại California hôm 17/3 ở Mỹ. Mẫu tin ngắn gọn nhưng súc tích xuất hiện trên mạng vốn vẹn 3 phút sau cơn chấn động. Có thể đưa tin một sự kiện chỉ trong tích tắc như vậy hẳn... "không phải người".

Quả thật, tác giả bản tin đó là một robot mang tên Quakebot!

## Câu chuyện về Quakebot

6 giờ 25 phút sáng thứ hai ngày 17/03/2014, Ken Schwencke, phóng viên tờ Los Angeles Times (cũng là người lập trình Quakebot), giật mình tỉnh giấc bởi cơn chấn động. Anh lập tức ra khỏi giường, đi thẳng đến máy tính đang mở sẵn. Đoạn văn ngắn mô tả những thông tin cơ bản nhất về trận động đất đã hoàn tất, đang nằm ngoan ngoãn chờ biên tập trong hệ thống. Đọc lướt bài viết, Ken nhấn

nút "xuất bản" và Los Angeles Times trở thành tờ báo đầu tiên đưa tin sự kiện ngày hôm đó. Toàn bộ quá trình từ lúc viết tin, chuyển biên tập đến khi xuất bản lên mạng chưa đầy 3 phút.

Nhanh chóng, chính xác và bất kể thời gian - đó là cách một writer-bot (robot phóng viên) như Quakebot làm việc. Hầu hết độc giả cho biết, họ không nhận thấy gì khác lạ khi đọc bản tin cho đến khi ô lên ngạc nhiên bởi câu cuối cùng: "**Bài viết được tạo**

**ra bởi một thuật toán của tác giả**" - ("This post was created by an algorithm written by the author").

Los Angeles Times không hẳn là tờ báo tiên phong. Thật bất ngờ khi biết rằng khoảng 350 triệu bài viết trên các trang thông tin bất động sản, đầu tư tài chính, bình luận thể thao... trong năm 2013 có tác giả là robot. Con số này được dự đoán tăng lên 1 tỷ trong năm 2014 và sẽ không dừng lại đó.



Tin tức đầu tiên về trận động đất 17-3 tại California được viết bởi một robot.



Phóng viên Ken Schwencke của Los Angeles Times là người lập trình Quakebot

A shallow magnitude 4.7 earthquake was reported Monday morning five miles from Westwood, California, according to the U.S. Geological Survey. The temblor occurred at 6:25 a.m. Pacific time at a depth of 5.0 miles. According to the USGS, the epicenter was six miles from Beverly Hills, California, seven miles from Universal City, California, seven miles from Santa Monica, California and 348 miles from Sacramento, California. In the past ten days, there have been no earthquakes magnitude 3.0 and greater centered nearby.

This information comes from the USGS Earthquake Notification Service and this post was created by an algorithm written by the author.

**(Bản tin do Quakebot thực hiện đăng trên Los Angeles Times ngày 17-3. Nếu không đọc câu cuối cùng, hiếm người nhận ra bản tin này không phải do người viết)**

## Bản dịch:

*Một trận động đất cường độ 4,7 xảy ra cách Westwood, California 5 dặm vào sáng thứ hai đã được báo cáo - Theo Cục Khảo sát Địa chất Mỹ (USGS). Cơn địa chấn xảy ra lúc 6 giờ 25 phút sáng (giờ Thái Bình Dương) ở độ sâu 5 dặm. Theo USGS, tâm chấn cách Beverly Hills 6 dặm, cách Universal City 7 dặm, Santa Monica 7 dặm và Sacramento 348 dặm. Trong 10 ngày qua, không xảy ra trận động đất nào có cường độ lớn hơn 3 ở gần đó.*

*Thông tin trên xuất phát từ dịch vụ thông báo của USGS và bài viết được tạo ra bởi một thuật toán của tác giả.*

### Những thuật toán viết văn

Robot nhà báo như “Quakebot” (thường gọi là robot writer, robot reporter hay robot journalist) hoàn toàn không phải những cỗ máy mô phỏng con người, hồi hải đến hiện trường phỏng vấn hay lóc cóc gõ bàn phím viết tin. Đó là những thuật toán được tạo ra bởi chương trình xử lý ngôn ngữ tự nhiên (Natural language processing - NLP), giúp máy tính phân tích và sắp xếp số liệu, từ ngữ, hình ảnh... một cách hệ thống thành văn bản có nghĩa. Có thể gọi nôm na là viết báo bằng thuật toán.

Thuật toán sẽ sàng lọc và tổng hợp thông tin nổi bật từ những nguồn chọn sẵn, rồi dựa vào các mẫu văn bản để sắp xếp câu, chữ, số liệu hình thành bài viết. Chẳng hạn, với dạng tin tức về động đất, người đọc thường quan tâm đến con số thương vong, tổn thất tài sản, nhận định của chuyên gia... Với bài tường thuật thể thao, những câu hỏi quan trọng cần trả lời là ai thắng? Tỷ số bao nhiêu? Chiến thắng lần thứ mấy? Số liệu nào liên quan?... Trường hợp Quakebot, thuật toán được thiết kế để thu thập thông tin từ Cục Khảo sát Địa chất Mỹ (The U.S. Geological Survey) và dựa vào các đoạn viết mẫu để tạo bản tin ngắn.

### Cuộc “xâm lăng” của binh đoàn robot

Nguồn lực công nghệ thông tin cực lớn đang tạo ra những trí thông minh nhân tạo ngang tầm, thậm chí bén nhọn hơn cả bộ não người. Mới vài năm trước, các bài viết của robot



*Xử lý ngôn ngữ tự nhiên (Natural language processing – NLP) là một nhánh nghiên cứu của trí thông minh nhân tạo, lập trình máy tính thực hiện các ứng dụng liên quan đến ngôn ngữ con người như: dịch tự động; nhận dạng chữ viết, giọng nói; tổng hợp tiếng nói; phân tích và xử lý văn bản; tìm kiếm thông tin...*

*Quá trình xử lý ngôn ngữ thường tiến hành qua các bước cơ bản:*

- Phân tích hình thái ngôn ngữ
- Phân tích cú pháp
- Phân tích ngữ nghĩa
- Tích hợp văn bản
- Phân tích thực nghĩa



**Google translate cho thấy sự tinh tế ngày càng được nâng cao của các công cụ NLP qua chất lượng các bản dịch.**

còn khá thô sơ, chủ yếu dùng trong các bản tin có cấu trúc thông dụng như tin thể thao, tin tai nạn, thông báo giá cả, bản tin mua bán... Nhưng hiện nay nhờ sự tham gia của các nhà báo vào khâu lập trình (như trường hợp phóng viên Ken Schwencke lập trình Quakebot), khả năng viết lách của các writer-bot ngày càng chuyên nghiệp. Có thể xử lý tốt cả loại tin, bài phức tạp, sử dụng nhiều nguồn tin kèm theo nhận định của chuyên gia.

Narrative Science có trụ sở tại Chicago là một trong những công ty chuyên sáng tạo công nghệ writer-bot đang rất ăn nên làm ra. Năm 2011, bài tường thuật chiến

thắng của đội bóng rổ Winconsin trước UNLV được một robot phóng viên của Narrative Science sáng tác trong vòng 30 giây. Trang tin tức lớn như Forbes hiện cũng sử dụng các writer-bot của Narrative Science. Tờ The Guardian thì đang thai nghén dự án lập trình đội phóng viên Guar-bot cho riêng mình.

Kristian Hammond, giám đốc công nghệ của Narrative Science dự báo, khoảng 90% tin tức có thể là do robot viết vào năm 2030 với chất lượng nâng cao đáng kể. Trong đó, năm 2017 được cho là thời điểm những robot nhà báo ồ ạt thâm nhập ngành công nghiệp tin tức.

### Nhà báo bị “lấn sân”?

Viết lách vốn được xem là nghề nghiệp đặc thù khó tự động hóa,

## Patriots Beaten by Broncos, 26-16, Fall in AFC Championship

Filed under Game Recap on January 19th, 2014

Sun, Jan 19 2014, 3:05 PM

Sports Authority Field at Mile High Stadium

 **Patriots 16** (12-4)  **Broncos 26** (13-3)

Team	1st	2nd	3rd	4th	Final
Patriots	0	3	0	13	16
Broncos	3	10	7	6	26

Recap | [Boxscore](#) | [Play-by-Play](#)

Share this recap

[Tweet](#) or [Like](#) One person likes this. Be the first of your friends.

Julian Edelman was a big receiving threat but could not reverse a tough defeat for the Patriots against the Broncos, 26-16 in Denver. With the loss, New England is knocked out of the playoffs.

Tom Brady accounted for two touchdowns (1 pass, 1 rush) on 24 of 38 passing for 277 yards and no interceptions. Peyton Manning also played well for the Broncos, and wound up 32 for 43 with 400 yards passing, two touchdowns and no interceptions.

Shane Vereen rushed for 34 yards on four carries and had five receptions for 59 yards. Edelman was a top performer, snagging 10 passes for 89 yards with a touchdown.

The New England defense allowed 107 yards rushing and 400 passing.

Broncos wide receiver Demaryius Thomas was impressive, racking up 134 receiving yards and one touchdown. Tight end Julius Thomas also put together a noteworthy performance with 85 receiving yards.

*Một số bản tin do robot viết: tin thể thao (trái) và tin mua bán (phải)*

## Dallas Metro Area's July First-Time Homebuyer Report

Posted by Homesnap on 7/30/2013 in Dallas Metro Area

The population of the Dallas Metro Area is very affluent compared to others in USA with a median income of \$58,465 per household. It's also considered an average metro area in terms of affordability with a cost of living adjustment score of 95.7. Its three largest counties are Dallas County, Tarrant County, and Collin County. The Dallas Metro Area is made up of homes that are of moderate age with a median of 25 years. Leading the list of top schools in the metro area are North Hills Primary School, North Hills School, and High School Campus.

### CULTURE

Recommended places to eat, according to Foursquare, in Dallas Metro Area include Meddlesome Moth, Whiskey Cake Kitchen & Bar, and Cafe Brazil. If you're looking to get a drink, a favorite is Round-Up Saloon. Lastly, The Dream Cafe is a well-known place to get some coffee.

### PRICE RANGE

The housing prices of the Dallas Metro Area are close to the country average with a median home value of \$129,600 and a current average list price of \$193,072. The Dallas Metro Area market showed continued weakness over the last 30 days, as sales fell 4.3% to 6,023 homes, the lowest amount in three months.

Curious as to which homes have been getting the most attention in the Dallas Metro Area? These listings have stirred up interest and produced the most views:



### TYPICAL DWELLING

The typical dwelling in the Dallas Metro Area is a three-bedroom detached home worth about \$173,592. To purchase a similar home, a buyer with good credit and a \$34,718 cash down payment could obtain a mortgage with a 4.53% interest rate and a \$1,123 monthly payment. The property tax

nhưng với sự phát triển ngày càng nhanh và mạnh của các writer-bot, nhiều nhà báo truyền thống đang e ngại bị máy móc dẫm chân vào công việc một ngày không xa.

Nghiên cứu của nhà khoa học Thụy Điển Christer Clerwall đầu năm 2014 cho thấy, rất khó để phân biệt bài báo do người viết hay robot viết. Trong các cuộc khảo sát, người đọc thường không nhận thấy nhiều sự khác biệt giữa bản tin của robot và bản tin do người viết. Thậm chí một số bài viết bởi robot còn được đánh giá nhiều thông tin và đáng tin cậy hơn so với bài viết của phóng viên.



*Kristian Hammond, giám đốc Narrative Science: "Trong 5 năm tới, một chương trình máy tính có thể thắng giải thưởng Pulitzer, và tôi sẽ bị nguyên rủa nếu đó không phải là công nghệ của chúng tôi".*

Giám đốc Hammond của Narrative Science quả quyết, biết đâu vào năm 2017, một chiếc máy tính cũng có thể tạo ra tác phẩm báo chí xứng đáng nhận giải Pulitzer.

### Hay được giải phóng?

Tuy vậy, theo ông Joe Procopio - Phó Tổng giám đốc của Automated Insights chuyên sản xuất các writer-bot tại Mỹ - robot được tạo ra hoàn toàn không thể thay thế nhà báo "bằng xương bằng thịt". Nhược điểm cố hữu của robot là không xác minh được thông tin thiếu khách quan nên đòi hỏi nguồn tin phải đáng tin cậy. Do đó, một số khách hàng của Automated Insights, trong đó có cả Yahoo và hãng truyền thông lớn Associated Press hiếm khi tiết lộ bài báo nào được tạo ra bởi thuật toán.

Công nghệ robot viết lách ra đời chỉ nhằm giải phóng phóng viên khỏi những công việc đơn điệu dễ gây nhàm chán như viết tin thông báo, quét tin hàng ngày... Máy

tính dễ dàng đảm trách phần việc đơn giản này với tốc độ và sự chính xác cao nhất dựa trên số liệu thu thập được. Nhờ đó người làm báo có thể toàn tâm toàn ý tập trung vào các mảng miếng thú vị hơn như phóng sự, điều tra,... Những tác phẩm báo chí thực thụ có chiều sâu và cảm xúc là điều mà các thuật toán khô khan không thể nào "hiểu" được.

Như trường hợp Quakebot, bản tin vẫn do robot tạo ra ban đầu chỉ có tác dụng "giữ chỗ" nhanh cho những bài viết dài hơi kế tiếp. Vài giờ sau khi bản tin của Quakebot xuất bản, công việc của các biên tập viên là triển khai thêm những khía cạnh có thể thu hút độc giả. Kết quả, một bài phân tích sâu và đầy đủ về cơn chấn động được hoàn thiện sau 71 lần cập nhật, bổ sung từ bản tin gốc của Quakebot.

Dẫn lời một chuyên gia ngành báo -John Lavine, Hiệu trưởng Học viện Báo chí Medill tại Northwestern- nhận xét về công nghệ robot phóng viên: "Tôi tin rằng, đây là loại công nghệ có thể giúp báo chí trở nên sâu sắc hơn"! □