

Nghề 'hot' nhất thế kỷ

✧ P. UYÊN

Tháng 7, hàng vạn sĩ tử bước vào kỳ thi tuyển sinh vào đại học, nơi sẽ trang bị cho họ một nghề để bước vào đời. Thường, với mong muốn chính đáng, nhiều thí sinh chọn nghề thời thượng, hy vọng ra trường để tìm được việc làm với mức lương "dễ thở". Có một nghề được dự báo sẽ rất "hot" trên thị trường lao động, nhưng ... hiện chưa có ở Việt Nam!

Khi Jonathan Goldman gia nhập vào tháng 6 năm 2006, LinkedIn (website kết nối chuyên gia) có số tài khoản đăng ký tham gia site gần 8 triệu, và con số này không ngừng tăng lên khi các thành viên tiếp tục mời bạn bè và đồng nghiệp tham gia. Nhưng dường như có khiếm khuyết gì đó khiến người dùng không thấy được sự kết nối giữa những người tham gia site này ở mức độ mà nhóm điều hành site mong đợi, "Nó giống như đi đến một buổi hội thảo và bạn thấy toàn người lạ, vì vậy bạn chỉ đứng trong góc nhâm nhi thức uống rồi có thể về sớm", theo một nhà quản lý site.

Dữ liệu site khi đó lộn xộn và khó phân tích, nhưng Goldman, tiến sĩ vật lý Đại học Stanford, nhìn thấy sự hấp dẫn từ hồ sơ phong phú của người dùng và khả năng kết nối các hồ sơ này. Goldman bắt đầu xây dựng những công thức kết nối, thử nghiệm các khả năng và tìm ra mô hình cho phép dự đoán các mạng lưới người dùng phù hợp với một hồ sơ nào đó. Goldman đề xuất các tính năng mới khai thác những tìm tòi trên để

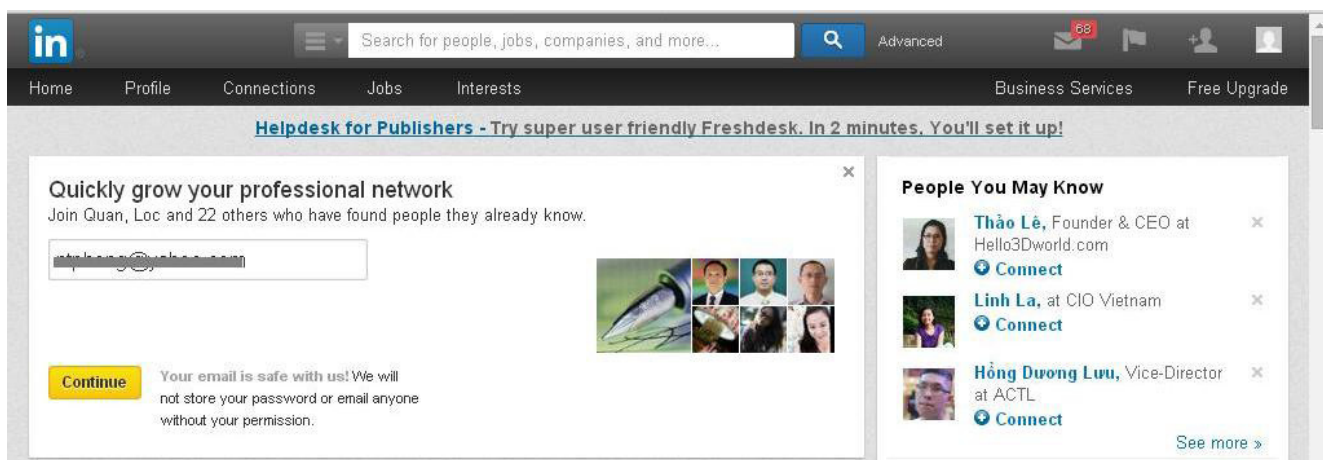
đem đến lợi ích cho người dùng, nhưng đội ngũ kỹ thuật của LinkedIn khi đó bị cuốn vào việc mở rộng quy mô site nên dường như không quan tâm. Thậm chí một số đồng nghiệp còn cười nhạo ý tưởng của Goldman. Site này đã có sẵn công cụ nhập danh bạ có thể truy ra hết các mối quan hệ của một thành viên, người dùng có thể tự tìm lấy, đầu cần LinkedIn tìm giúp các cộng đồng phù hợp với họ.

Thật may, Reid Hoffman, đồng sáng lập và là giám đốc điều hành của LinkedIn lúc đó (giờ là chủ tịch hội đồng quản trị), có niềm tin vào sức mạnh của việc phân tích thông tin qua kinh nghiệm tại PayPal, và ông đã cho Goldman quyền tự quyết. Cụ thể, ông cho phép Goldman đưa ra các mô-đun nhỏ dưới hình thức quảng cáo trên các trang phổ biến nhất của LinkedIn (việc này phá vỡ quy trình tung ra sản phẩm của LinkedIn).

Thông qua các mô-đun này, Goldman thử nghiệm phản ứng của người dùng

khi hiển thị tên của những người họ chưa kết nối nhưng có khả năng quen biết, chẳng hạn như những người học cùng trường hay làm việc cùng công ty. Ông thực hiện bằng cách quảng cáo "môi" hiển thị ba mối quan hệ tốt nhất dựa trên hồ sơ cá nhân của người dùng. Kết quả thật ấn tượng, tỷ lệ nhấp chuột vào các quảng cáo này cao chưa từng thấy. Goldman tiếp tục hoàn thiện công thức sinh ra các kết nối để nghị kết hợp với các ý tưởng nối mạng như "liên hệ tam giác" (nếu bạn biết anh A và chị B, nhiều khả năng A và B biết nhau).

Không mất nhiều thời gian để các nhà quản lý hàng đầu của LinkedIn nhận ra ý tưởng này tốt và cho phép nó trở thành chức năng chính thức. Và mọi thứ vào guồng. Quảng cáo "People You May Know" (người mà bạn có thể biết) tạo ra hàng triệu lượt xem, đạt tỷ lệ nhấp chuột cao hơn 30% so với các quảng cáo mời gọi khác trên LinkedIn. Nhờ tính năng này, đường tăng trưởng của LinkedIn dốc lên đáng kể.



Nghề mới thời “dữ liệu lớn”

Goldman là ví dụ điển hình của một “vai” mới quan trọng trong các tổ chức, công ty: “nhà khoa học dữ liệu”, chức danh này được DJ Patil và Jeff Hammerbacher đưa ra vào năm 2008, để mô tả vị trí công việc quản lý và phân tích dữ liệu tại LinkedIn và Facebook. Cách đây 5 năm không hề có quảng cáo tuyển dụng chuyên gia về dữ liệu. Giờ, đây là công việc “hot” nhất ở Thung lũng Silicon (Mỹ), và hàng ngàn nhà khoa học dữ liệu hiện đang làm việc tại các công ty lớn nhỏ. Sự xuất hiện của họ phản ánh bối cảnh hiện nay các công ty đang phải vật lộn với thông tin đến từ nhiều nguồn (như Facebook, điện thoại di động hay cảm biến) với khối lượng lớn chưa từng có.

“Dữ liệu lớn” (Big Data), giờ hầu hết mọi người đều nghe nói đây là “thứ lớn lao kế tiếp” sẽ tạo nên cuộc cách mạng thay đổi cách chúng ta làm việc, sinh sống và giao tiếp. Nhưng ai sẽ quản lý cơ sở dữ liệu Walmart có hơn 2,5 petabyte (tỷ tỷ byte) dữ liệu từ 1 triệu giao dịch khách hàng mỗi giờ? Ai tại YouTube thực hiện việc phân tích 48 giờ video được tải lên trang web này mỗi phút? Phần lớn sự háo hức với dữ liệu lớn hiện nay chỉ chú trọng đến các công nghệ “thuần phục” dữ liệu, trong khi khâu quan trọng không kém chính là những con người có kỹ năng khai thác dữ liệu. Về mặt này, cầu đang vượt cung.

Theo một nghiên cứu của McKinsey, phân tích dữ liệu lớn có thể giúp các nhà bán lẻ tăng biên độ lợi nhuận hoạt động lên 60%, ngành công nghiệp chăm sóc sức khỏe toàn cầu có thể giảm 8% (hay 200 tỷ USD) chi phí hàng năm. Nghiên cứu này cũng cảnh báo sự thiếu hụt khoảng 140.000 đến 190.000 lao động có kỹ năng phân tích dữ liệu vào năm 2018.

Sử dụng các nhà khoa học dữ liệu gồm các công ty tên tuổi như Microsoft, Amazon và Google, nhưng công việc này không chỉ dành cho các tập đoàn công nghệ lớn. Cơ quan chính phủ, công ty dịch vụ công ích, tổ chức tài

chính, bệnh viện, công ty sản xuất, bán lẻ và các công ty quảng cáo cũng rất cần có loại lao động “thời thượng” này.

Trở thành nhà khoa học dữ liệu

Các nhà khoa học dữ liệu là những ảo thuật gia của thời đại dữ liệu lớn. Họ nghiên cứu dữ liệu, sử dụng mô hình toán học để phân tích và diễn giải, sau đó đề nghị cách thức sử dụng thông tin để đưa ra quyết định.

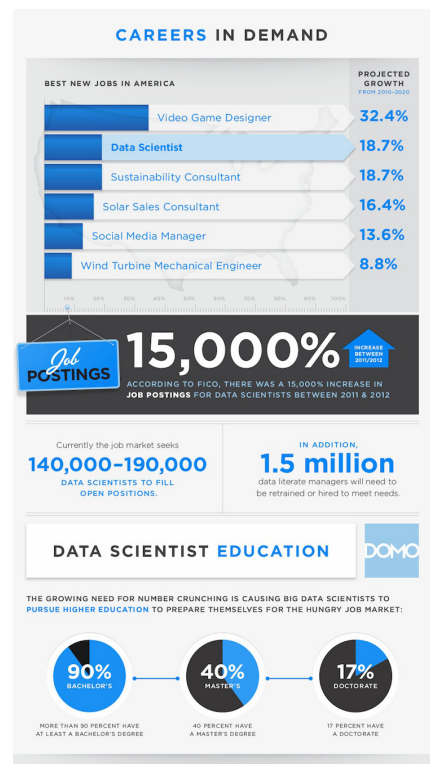
Tin tốt là trong vài năm qua xuất hiện nhiều chương trình đào tạo (dưới nhiều hình thức khác nhau) đáp ứng nhu cầu chuyên gia về dữ liệu lớn. Ví dụ như các chương trình khoa học dữ liệu chuyên sâu hay phân tích nâng cao tại Viện Nghiên cứu Khoa học và Kỹ thuật Dữ liệu Columbia, Đại học Berkeley, Đại học Carnegie Mellon, Viện Công nghệ Illinois, Đại học Imperial, Đại học North Carolina, Đại học Syracuse và Đại học Tennessee. Các trường khác có giảng dạy khoa học dữ liệu như Đại học New York, Stanford, Northwestern, George Mason, California và Indiana.

Năm rồi Đại học Columbia đưa ra các chương trình cao học và chứng chỉ chuyên về dữ liệu, lớp thạc sĩ về phân tích tại Đại học San Francisco thì sắp tốt nghiệp. Cloudera, công ty bán phần mềm xử lý dữ liệu lớn, công bố sẽ làm việc với 7 trường đại học để huấn luyện cho sinh viên sắp tốt nghiệp về các công nghệ dữ liệu lớn.

Trong khi đó, các công ty như IBM đang hợp tác với các trường đại học để đào tạo kỹ năng về dữ liệu lớn. Và những người tiên phong, các chuyên gia khoa học dữ liệu đầu tiên của Yahoo và LinkedIn giờ rải rác khắp giới công nghệ, tự nguyện đào tạo và truyền cảm hứng cho thế hệ các nhà khoa học dữ liệu kế tiếp. Chương trình học bổng Khoa học dữ liệu (Insight Data Science Fellowship Program) do Jake Klamka khởi xướng là ví dụ điển hình. Học bổng đào tạo sau tiến sĩ chuyên sâu 6 tuần này lấp khoảng trống giữa kiến thức hàn lâm và thực tế trong lĩnh vực khoa học dữ liệu và phân tích. □



Nếu bạn có thiên hướng toán học và dữ liệu, và đang tìm kiếm một công việc “tuyệt vời”, không có lựa chọn nào tốt hơn ngành khoa học dữ liệu. Do sự bùng nổ dữ liệu, các công ty hiện có thông tin (dữ liệu) nhiều chưa từng thấy và không có đủ nhân lực để giải mã hết đồng dữ liệu này. Đây là thị trường lớn cho các nhà phân tích định lượng và những người có khả năng “vẽ viễn cảnh từ đồng dữ liệu đồ sộ”.



Nguồn: VentureBeat.

Theo Anjul Bhambhri, phó chủ tịch về sản phẩm dữ liệu lớn của IBM, nhà khoa học dữ liệu là sự kết hợp “chuyên gia phân tích và nghệ sĩ”, họ “có tính tò mò, có thể nhìn dữ liệu và vẽ nên đường nét (xu hướng)”.