AI-01 · Bài 2

Mỗi câu hỏi cần một loại dữ liệu riêng, và lấy nhầm loại thì câu trả lời vẫn trông chắc chắn

Lớp Dữ liệu nào dùng được trong phát triển sản phẩm

m.school by MES LAB

Học xong bài này bạn sẽ

- viết được câu hỏi của mình ra giấy trước khi đi tìm dữ liệu, tới mức biết câu trả lời sẽ có hình dạng gì.
- Chỉ ra được với một câu hỏi cụ thể thì bộ dữ liệu đang có trong tay trả lời được phần nào và bỏ trống phần nào.
- Nói được vì sao thêm thật nhiều dòng số cũng không cứu được một câu hỏi đang hỏi nhầm loại dữ liệu.

1 · Viết câu hỏi ra giấy trước, rồi mới đi tìm dữ liệu

- Phần lớn các cuộc đi tìm dữ liệu bắt đầu bằng một câu như «xem thử khách hàng nghĩ gì về sản phẩm của mình».
- Một câu hỏi dùng được có ba phần.
- Thử sửa câu trên theo ba phần đó.
- Có một phép thử nhanh để biết câu hỏi đã đủ rõ chưa: nói trước xem câu trả lời sẽ có hình dạng gì.
- Còn một câu nữa nên hỏi trước khi bắt đầu: ai sẽ dùng câu trả lời này, và để quyết định việc gì.

2 · Hai con số cùng mang tên doanh số có thể trả lời hai câu hỏi khác nhau

- Hàng từ nhà máy đi tới người dùng cuối qua hai bước, và mỗi bước cho một con số.
- Bước thứ nhất là hàng rời nhà máy để vào kho của đại lý, siêu thị hay sàn bán hàng.
- Trong sổ sách, cả hai đều có thể được gọi là doanh số.
- Cái bẫy nằm ở chỗ con số thứ nhất dễ đo hơn nhiều.
- Chỗ này dẫn tới một tình huống quen thuộc: hàng vào kênh đều đặn, báo cáo tháng nào cũng đạt, trong khi hàng nằm trên kệ không đi đâu cả.
- Bài học chung rộng hơn chuyện bán hàng: trước khi dùng một con số, hãy hỏi nó được đo ở điểm nào trên đường đi của sự việc.

3 · Dữ liệu ghi lại việc đã xảy ra, nên nó nói được chuyện gì xảy ra mà không tự nói vì sao

- Dữ liệu là dấu vết của việc đã xảy ra.
- Giới hạn thứ nhất: dữ liệu cho thấy hai việc đi cùng nhau, chứ không nói việc này gây ra việc kia
- Vì vậy khi dữ liệu báo có chuyện không ổn, việc nên làm không phải là kết luận ngay, mà là hỏi: trong cùng khoảng thời gian ấy, còn những gì đã đổi.
- Giới hạn thứ hai: không có dữ liệu nào về thứ chưa từng tồn tại
- Với loại câu hỏi ấy, đường đi không phải là tìm thêm dữ liệu cũ, mà là tạo ra dữ liệu mới: làm một bản mẫu thô, đưa cho người dùng thật, và ghi lại họ làm gì với nó.

4 · Câu hỏi quan trọng thì đừng dựa vào một nguồn, và chỗ các nguồn lệch nhau là chỗ đáng xem

- Cách thu nào cũng bỏ sót một phần của sự việc, và phần bị bỏ sót thì nằm ngoài tầm nhìn của chính cách thu đó.
- Cách xử lý là ghép nhiều nguồn có điểm mù khác nhau, rồi xem chúng có chỉ về cùng một chỗ không.
- Chữ độc lập là chữ then chốt.
- Và đây là phần dễ bị bỏ qua nhất: khi các nguồn lệch nhau, đừng chọn nguồn hợp ý mình.
- Ba nguồn độc lập cùng chỉ về một chỗ, và chỗ hai nguồn lệch nhau là chỗ đáng đi xem tiếp. -->

Bảng tổng hợp

Bảng tổng hợp: câu hỏi nào thì đi với loại dữ liệu nào

Câu hỏi bạn đang có	Loại dữ liệu trả lời được	Loại trông giống mà không trả lời được
Có bao nhiêu người gặp chuyện này, và chuyện nào nhiều hơn chuyện nào	hồ sơ đếm được: phiếu bảo hành, lý do đổi trả, số cuộc gọi phân theo loại	vài câu chuyện kể lại, dù câu nào cũng rõ
Vì sao người ta làm như thế	lời kể ghi nguyên văn, quan sát tại chỗ, biên bản khiếu nại	bảng số, vì nó cho thấy có chuyện mà không cho biết hoàn cảnh
Sản phẩm mới có được người mua đón nhận không	hàng bán tới tay người dùng, lượng tồn ở kênh, số đại lý tự đặt thêm	hàng đẩy vào kênh phân phối, vì nó dừng ở nửa đường

Bảng tổng hợp: câu hỏi nào thì đi với loại dữ liệu nào tiếp

Câu hỏi bạn đang có	Loại dữ liệu trả lời được	Loại trông giống mà không trả lời được
Thiết kế đã đạt mức đặt ra chưa	các phép đo và phép thử chính nhóm làm trên bản mẫu, kèm chỗ mẫu hỏng	phản hồi của người dùng về bản đang bán, vì đó là bản khác
Người ta có mua thứ chưa ai từng bán không	dữ liệu mới tạo ra bằng bản mẫu đưa cho người dùng thử	mọi dữ liệu bán hàng quá khứ, vì nó chỉ ghi thứ đã được bày bán

Phần giá trị nhất của bài

Chỗ người ta hay làm sai

- **Đi tìm dữ liệu trước khi viết câu hỏi ra giấy**

  Cách tránh là viết câu hỏi ra trước, rồi mới mở dữ liệu, và nếu giữa chừng câu hỏi đổi thì ghi lại là nó đã đổi.

- **Lấy dữ liệu dễ lấy thay cho dữ liệu đúng**

  Cách tránh là ghi thẳng vào báo cáo một dòng nói rõ dữ liệu này nói về ai và bỏ sót ai, đặt ngay cạnh kết luận chứ không giấu ở phần phụ lục.

- **Hỏi dữ liệu cũ một câu về thứ chưa từng tồn tại**

  Cách tránh là nhận ra sớm rằng câu hỏi này thuộc loại phải tạo dữ liệu mới, và bỏ một phần nhỏ công sức đó vào việc làm một bản mẫu thô.

Ví dụ để khái niệm sáng ra

Ví dụ minh hoạ

- Năm 2015, hai tác giả làm việc tại Netflix công bố trên một tạp chí khoa học bài viết về hệ thống gợi ý phim của hãng.
- Khi chuyển sang chiếu trực tuyến, hãng có một loại dữ liệu khác: mỗi người xem gì, xem vào lúc nào, xem trên thiết bị gì, tìm thấy phim ở chỗ nào trong sản phẩm...

Điều rút ra: câu hỏi đổi thì loại dữ liệu trả lời được nó cũng đổi. Điểm sao không sai, nó chỉ đang trả lời một câu hỏi khác.

Ví dụ dưới đây lấy từ một bài báo khoa học công khai, bạn tra lại được.

Bài tập tự chấm

Thử nhìn lại việc của mình

- [ ] Lấy một câu hỏi đang thật sự chặn một quyết định trong công việc của bạn.
- [ ] Viết lại nó cho đủ ba phần ở mục 1: đối tượng, điều muốn biết, khoảng thời gian.
- [ ] Viết thêm một dòng nói câu trả lời sẽ có hình dạng gì.
- [ ] Rồi kẻ ba cột: bộ dữ liệu bạn định dùng, phần câu hỏi mà nó trả lời được, phần câu hỏi mà nó không chạm tới.
- [ ] Cuối cùng, tìm một nguồn thứ hai độc lập với nguồn thứ nhất và ghi nó vào.

Cách tự chấm và lời giải mẫu nằm trong bài trên trang web.

m.school

Đọc trọn bài và làm bài tập tại

mschool.meslab.org/lop/AI-01/bai-02.html

m.school by MES LAB · học không mất phí, không cần đăng nhập