

# [Spec] Stopwords

## Tổng quan:

Khi khách hàng enable tính năng Instant Search Widget, mỗi khi KH input thì app sẽ tự động lọc ra những từ ngữ không có nhiều ý nghĩa trong query để giúp hiển thị kết quả chính xác hơn

### VD: Case study

- Merchant bán đồ bơi
- KH input query: **a black swimsuit for man with flower pattern**
- Admin site có stopwords '**a**', '**for**', '**with**'

-> query được đưa vào tìm kiếm: **black swimsuit man flower pattern**

## Source index

Có 2 nguồn lấy stopwords

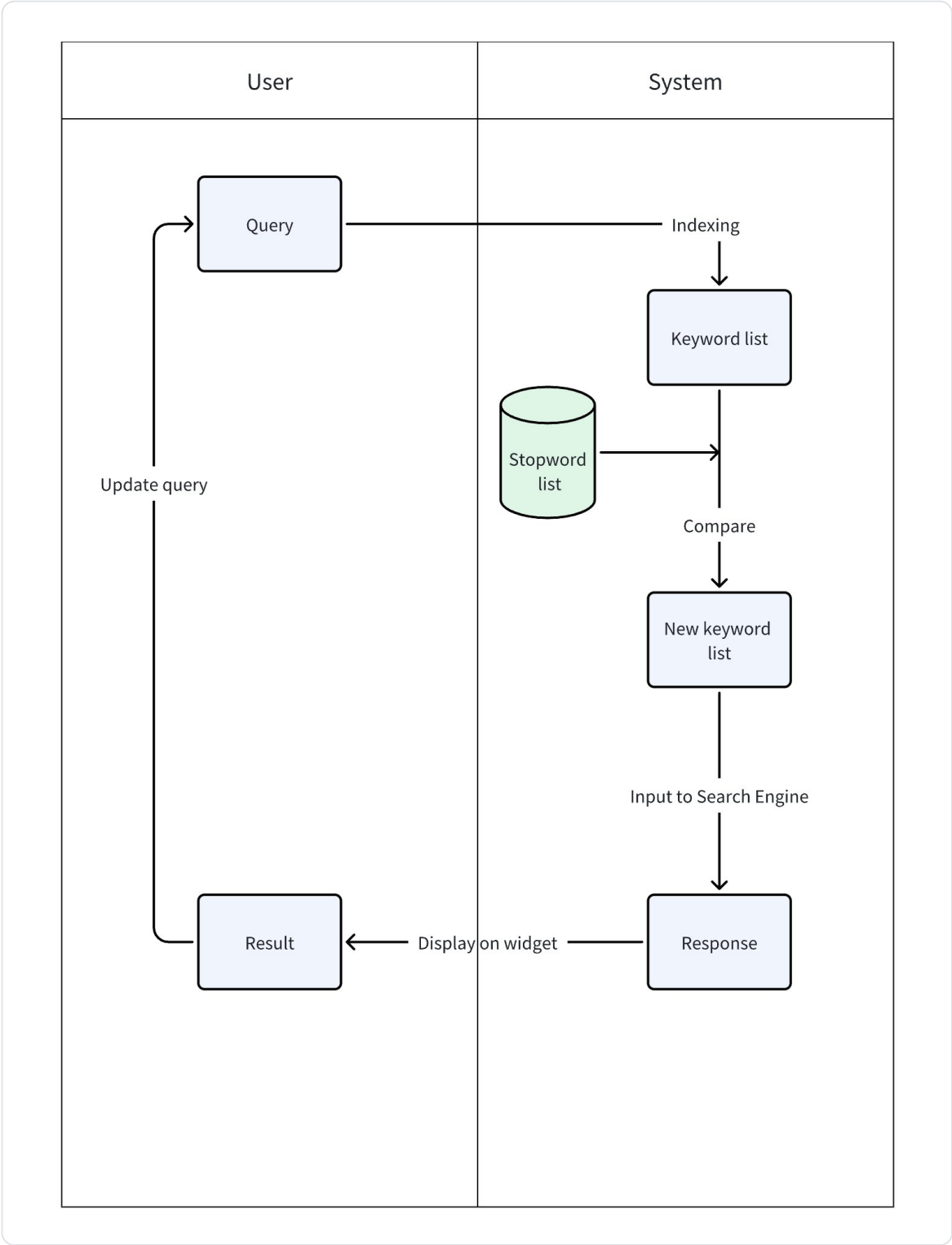
- **Nguồn tổng hợp của app:** Source stopwords do bên app tự thu thập và apply vào store của khách hàng, nguồn stopwords này sẽ luôn luôn được sử dụng và không hiển thị trong admin site của khách hàng
- **User input manually:** Nguồn này được hiển thị trên admin site của KH, do khách hàng nhập vào bằng tay hoặc bulk import qua csv file
  - Chú ý: Khi KH mới cài app, sẽ có 1 list các stopwords được add tự động trước, list bao gồm:
    - a
    - am
    - an
    - as
    - and
    - at
    - be
    - been
    - did

- didn' t
- do
- does
- doesn' t
- don' t
- had
- hadn' t
- has
- hasn' t
- have
- haven' t
- if
- is
- isn' t
- of
- on
- or
- should
- shouldn' t
- so
- the
- to
- was
- wasn' t
- were

## Dataflow

- Step 1: User input query
- Step 2: query được indexing thành các keyword
- Step 3: Sau khi finish indexing, list keyword sẽ được so sánh với list stopwords hiện có. Những keyword trùng với stopwords list sẽ được loại bỏ

- Step 4: List keyword được đưa vào search engine để thực hiện phân tích và trả về kết quả
- Step 5: User update query (nhập thêm kí tự) -> thực hiện lại từ Step 2



Figma:

<https://www.figma.com/file/dHTNp9Dd43Sbzy0hNWGMUC/Searching?type=design&node-id=1%3A6&mode=design&t=IGzVjswUVK14vx7B-1>