

CHAT HƯỚNG DẪN XÂY DỰNG KNOWLEDGE BASE

==

Hiểu rồi 🤖. Ý bạn là muốn mình **mở rộng bộ Knowledge Base JSON lên mức đầy đủ gấp nhiều lần (~1000 lần so với bản mẫu)** để có thể bao quát hầu hết tình huống tư vấn về cân nặng, béo phì và chế độ ăn uống.

Với mức độ “gấp 1000 lần”, ta sẽ không thể viết một JSON cực lớn trong một lần (nó sẽ dài đến hàng chục MB). Thay vào đó, mình sẽ hướng dẫn bạn **cách thiết kế và sinh tự động** Knowledge Base khổng lồ này từ một **schema chuẩn**.

📌 Cách mở rộng Knowledge Base

Thay vì chỉ có vài chục entry (BMI, vòng eo, lifestyle...), bạn có thể tạo ra **hàng chục ngàn entries** nhờ các nhóm sau:

1. Phân loại chi tiết

- BMI: thay vì chỉ 6 mức, chia nhỏ thành từng bước 0.1 → hơn 200 entries.
- Vòng eo: chia theo từng nhóm tuổi & giới tính → hàng trăm entries.
- Bệnh nền đi kèm (tiểu đường, huyết áp, cholesterol, bệnh gan...) → mỗi bệnh thêm rule riêng.

2. Thực phẩm & dinh dưỡng

- Tạo bảng dữ liệu chi tiết 5000+ loại thực phẩm (từ USDA hoặc FAO).
- Mỗi entry: calories, protein, carbs, fat, fiber, glycemic_index, advice.
- Ví dụ: "ăn 100g gạo lứt → cung cấp 111 calo, GI thấp, phù hợp người thừa cân."

3. Thực đơn (meal plan)

- Sinh ra hàng ngàn mẫu **meal plan** theo từng:
 - Mục tiêu (giảm cân, tăng cân, duy trì)
 - Chế độ ăn (Địa Trung Hải, DASH, Keto, Low-carb, Vegetarian...)
 - Lượng calo mục tiêu (1200, 1500, 1800, 2000, ...)

4. Vận động

- Database bài tập thể dục:
 - Cardio (chạy bộ, bơi, đạp xe, HIIT...)

- Sức mạnh (squat, deadlift, push-up...)
- Yoga, giãn cơ, thiền
- Mỗi bài: calories_burned_per_30min, difficulty, suitable_for_BMI_range.

5. Rules (luật suy diễn)

- Kết hợp điều kiện: BMI + tuổi + giới + bệnh nền + mức vận động → tạo ra **rule-based tư vấn**.
- Ví dụ:


```
{
  "if": {
    "BMI": ">=30",
    "age": "40-60",
    "disease": "diabetes"
  },
  "then": {
    "risk": "Nguy cơ rất cao biến chứng tim mạch",
    "advice": [
      "Giảm carb < 100g/ngày",
      "Ưu tiên rau lá xanh, cá biển, thịt trắng",
      "Khám sức khỏe định kỳ mỗi 3 tháng"
    ]
  }
}
```

Schema JSON mở rộng

Bạn có thể tổ chức như sau (chỉ demo cấu trúc, khi nhân bản có thể lên đến hàng chục ngàn entries):

```
{
  "BMI_Classification": [...200+ entries...],
  "Waist_Risk": {...chia theo tuổi, giới...},
  "Diseases": [
    {
      "name": "Diabetes Type 2",
      "risk_factors": ["BMI >= 25", "waist > threshold", "family_history"],
      "diet_recommendations": ["Hạn chế đường, tinh bột nhanh", "Ăn ngũ cốc nguyên hạt"],
    }
  ]
}
```

```

"exercise_recommendations": ["Đi bộ nhanh 30 phút/ngày", "Tập tạ nhẹ 2-3 buổi/tuần"]
},
{
"name": "Hypertension",
"diet_recommendations": ["Chế độ ăn DASH", "Hạn chế muối < 5g/ngày"],
"exercise_recommendations": ["Yoga, thở sâu", "Cardio cường độ vừa"]
}
],
"Foods": [...5000+ items...],
"Meal_Plans": [...hàng chục nghìn thực đơn...],
"Exercise": [...200+ bài tập với chi tiết...],
"Rules": [...10,000+ rule kết hợp...]
}

```

Cách sinh dữ liệu khổng lồ

1. Nguồn mở dữ liệu thực phẩm:

- USDA FoodData Central (\approx 5000+ loại thực phẩm)
- FAO food composition tables

2. Sinh tự động bằng script (Python/Node.js):

- Sinh ra các entries từ công thức (ví dụ BMI từ 15 \rightarrow 45 chia nhỏ từng 0.1).
- Sinh meal plan bằng cách kết hợp ngẫu nhiên thực phẩm theo macro mục tiêu.
- Sinh rule dựa trên tổ hợp BMI + disease + activity_level.

3. Lưu trữ:

- JSON khổng lồ (50-100MB) nếu chỉ lưu knowledge base.
- Hoặc **database** (MongoDB/PostgreSQL) để dễ query.

 Mình có thể viết cho bạn một **script Python sinh tự động Knowledge Base JSON đầy đủ (10.000+ entries)**.

Bạn có muốn mình làm **script Python để sinh file JSON** đó không, hay bạn chỉ cần mình **soạn một bộ JSON “full” vài trăm entries mẫu** để bạn hình dung trước?