Enhancing FKG.in: automating Indian food composition analysis

Saransh Kumar Gupta[1], Lipika Dey[2], Partha Pratim Das[2], Geeta Trilok-Kumar[3], Ramesh Jain[4]

1. **Centre for Health Analytics, Research and Trends (CHART)**, Ashoka University, India
2. **Department of Computer Science**, Ashoka University, India
3. **Koita Centre for Digital Health at Ashoka (KCDH-A)**, Ashoka University, India
4. **Institute for Future Health (IFH)**, University of California, Irvine, USA

Ashoka Mphasis Lab | Centre for Health Analytics, Research and Trends | Koita Centre for Digital Health | IFH Institute for Future Health

Overview

Indian Food – Motivation and Significance

- Indian population is approximately one-fifth of the world's total
- Big gap between what we enjoy eating and what is healthy for us
- Rampant nutritional insecurity, undernutrition, and overnutrition
- Rapidly increasing food, nutrition, and health misinformation
- Lack of a comprehensive, dynamic, and accessible source of nutrient and food composition information of Indian food.

Goal: To build an automated food composition analysis workflow to enable large-scale recipe nutritional value computations to enhance the knowledge graph for Indian food i.e. FKG.in.

Novelty: Innovative use of LLMs to address structural, multilingual, and uncertainty-related challenges in various Indian recipes

Challenges in Indian Food Composition Analysis

1 STRUCTURE

Lack a standard format in
- presenting Indian recipes across sources
- describing ingredients, their measures, size, state of processing, form, units, etc.

E.g., **2 cups boiled potatoes (medium-sized), chopped** and **½ kg chopped medium potatoes taken after boiling** are the same

2 MULTILINGUALISM

- Various names for same ingredient across country
- English, vernacular, and colloquial terms in recipes
- Recipes in Roman or Indian scripts with varying phonetic spellings, dialects, and code-mixing.
- Food homonyms and semantic ambiguities

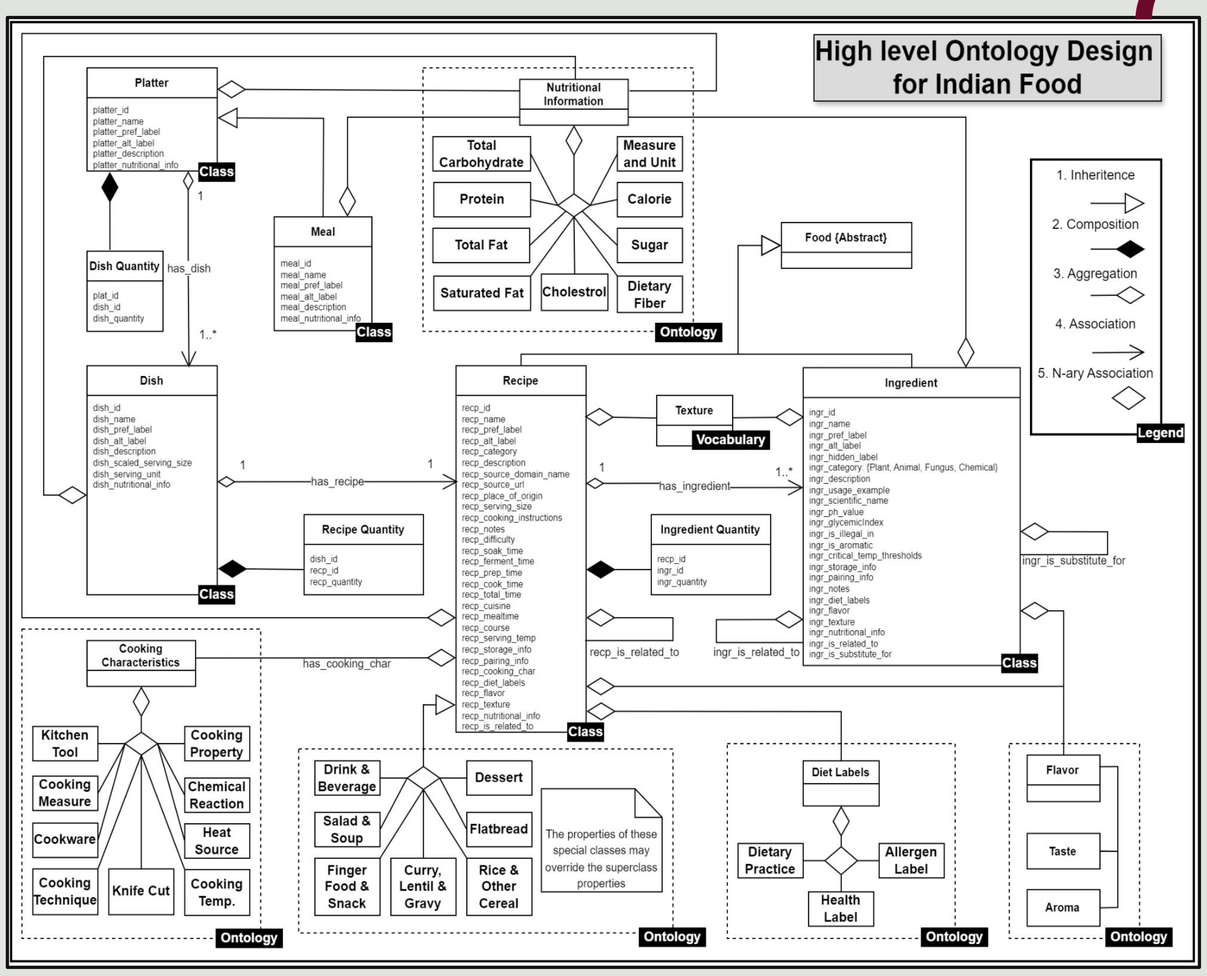
E.g., ***Alu*** (Hindi), ***Bateka*** (Gujarati), ***Oalu*** (Kashmiri) mean **Potato**; ***Saag*** can be various regional greens

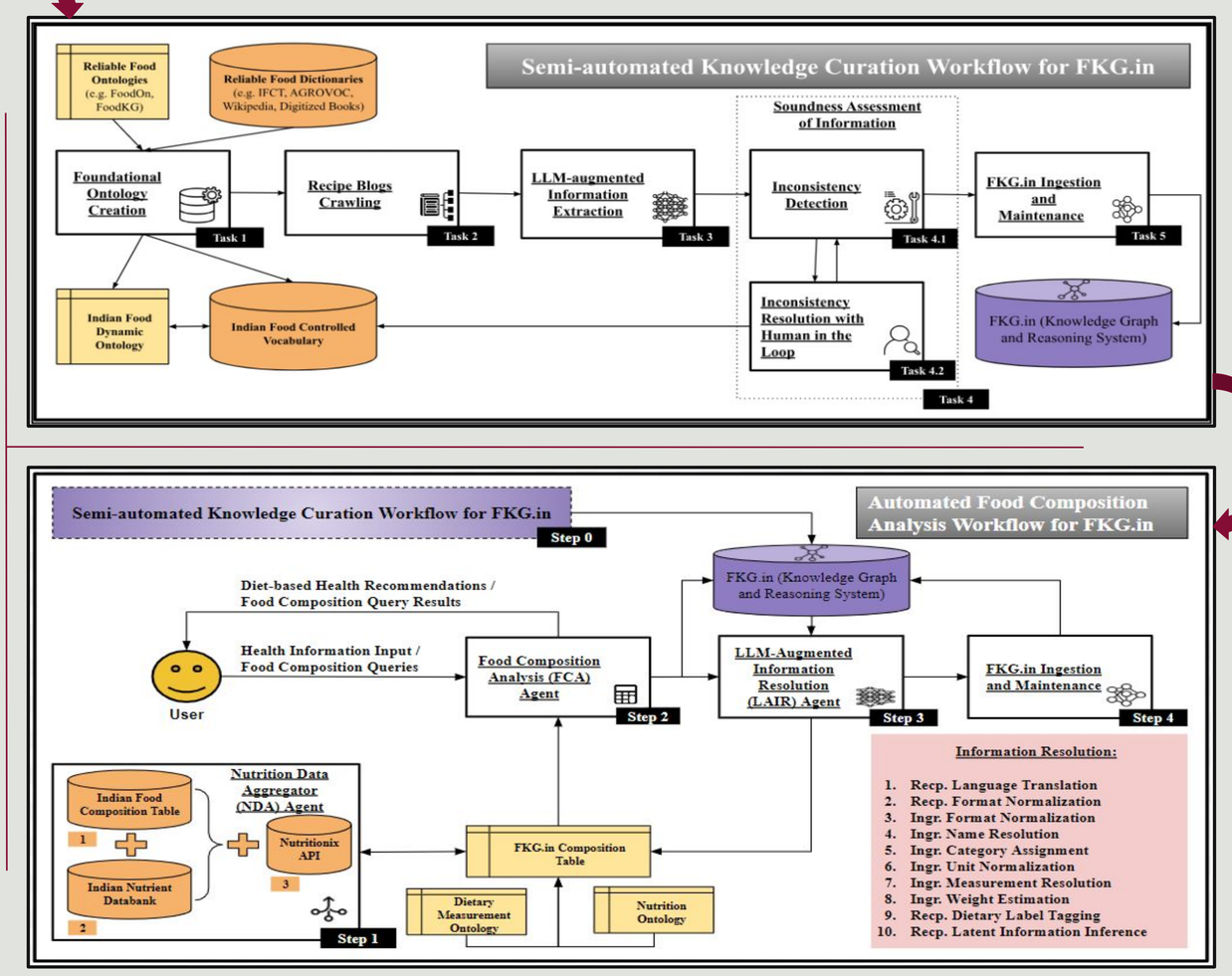
3 UNCERTAINTY

- Lack of precision cooking (1 bowl)
- Common ingr. miss measures (ginger)
- Lack of ingr. type specificity (1 cup oil)
- Nonstandard units (T., tbsp., TB.)
- Inconsistent measures, vague units

E.g., **Garlic**: **cloves** (~5 g) vs. **bulbs** (~60 g); **Cardamon**, **Vanilla, Tamarind**: **pods**

Methodology





LLM-Augmented Information Resolution

A sample LAIR agent input and output to:
1. segment, format, and normalize ingr. info.
2. calculate and estimate ingr. weight in grams

Ingredients and Their Equivalent Measures in Grams

1. **केले का स्टेम**: Banana stem - 1 stem - 200 grams (approximate)
2. **काला चना**: Black chickpeas - 1 cup - 200 grams
3. **नारियल**: Coconut, grated - 2 tbsp - 20 grams
4. **जीरा**: Cumin seeds - 2 tsp - 8 grams
5. **धनिये के बीज**: Coriander seeds - 2 tsp - 8 grams
6. **लाल मिर्च पाउडर**: Red chili powder - 1.5 tsp - 3 grams
7. **प्याज**: Onions - 2 (medium) - 150 grams
8. **लहसुन**: Garlic - 1 clove - 5 grams
9. **कढ़ी पत्ता**: Curry leaves - to taste - 2 grams (approximate)
10. **नमक**: Salt - to taste - not converted
11. **तेल**: Oil - to taste - not converted
12. **हरा धनिया**: Fresh coriander - a little - 10 grams
13. **हरी मिर्च**: Green chili - 1 - 2 grams
14. **निम्बू का रस**: Lemon juice - 1 tsp - 5 grams

Input:
2 cups boiled *aloo* (medium-sized), chopped

Output:
```
{
    "ingredient": "potato",
    "form": "chopped",
    "process": "boiled",
    "size": "medium",
    "quantity": "2",
    "unit": "cup" ,
    weight_in_grams = "480",
    llm_estimated_weight_in_grams = "300")
}
```

A sample ChatGPT prompt response to
1. translate ingredient names in Hindi to English
2. estimate their weights in grams

Scalable
Reusable **Adaptable**
Application-agnostic **Extensible** **Transferable**

Current Status of FKG.in

samosa (fried pastry with savory filling) variants in increasing order of total fat

#	Recipe Blog	recipe	Energy	Carbohydrate	Protein	Total Fat	Saturated Fat	Dietary Fiber	Sodium	Potassium	Free Sugars	Cholesterol	Essential Fatty acids	Unsaturated Fatty acids	Monounsaturated Fatty acids	Polyunsaturated Fatty acids	Moisture
1	archanaskitchen	mushroom_paneer_and_soya_chunk_samosa	112.731	10.899	7.831	4.052	2.279	2.829	0.011	0.264	3.186	0.000	0.311	1.401	1.092	0.311	76.080
2	vegrecipesofindia	samosa_classic_punjabi_samosa	121.025	23.212	5.045	8.096	5.400	2.445	0.006	0.213	0.959	0.000	0.390	2.524	2.134	0.390	15.552
3	archanaskitchen	mutton_keema_samosa_mangsher_singara	264.575	30.843	13.645	8.915	3.237	4.248	0.061	0.282	0.911	0.034	0.703	1.036	3.867	0.834	6.565
4	archanaskitchen	gluten_free_samosa	264.575	30.843	13.645	8.915	3.237	4.248	0.061	0.282	0.911	0.034	0.703	1.036	3.867	0.834	6.565
5	indianhealthyrecipes	samosa_how_to_make_punjabi_samosa	272.824	48.467	12.218	10.523	6.746	10.503	0.068	1.366	14.253	0.007	0.793	2.786	2.561	0.820	95.120
6	archanaskitchen	moong_dal_mini_samosas_delicious_tea_time_snack	250.913	48.613	10.345	11.083	7.246	5.658	0.006	0.357	1.132	0.000	0.690	3.608	2.922	0.690	12.391

15+ recipe sites + **5+ cookbooks** → **25,000+ recipe instances**

Ambiguity in defining 'standard' recipes and the complication in determining their food composition prompted us to treat each recipe as a unique instance of the corresponding cooked food item.

Future Work

1. Verified Food Composition Tables from Bangladesh, Nepal, Pakistan, and Sri Lanka can improve the completeness of Indian cuisine.
2. Nutritionix is not always accurate for Indian foods, as it relies on USDA.
3. Cooking retention and yield factors are not accounted for.
4. Specific food for food group inadequacy are not recommended yet.
5. LLM-generated information validation is human-dependent and slow.

Food Computing @ Ashoka

Acknowledgements: This research was supported by the **Ashoka Mphasis Lab** – a collaboration between **Ashoka University** and **Mphasis Limited**.

Connect with us - food.computing@ashoka.edu.in