

## Challenges encountered – food allergens

The following is a sampling of some of the challenges we encountered trying to work with the Wikipedia data as well as some of the ChatGPT output.

### Example:

| name = Hao Kuih (𩺰粿)/Hou Guo

| main\_ingredient = Flour and sweet potato flour, seafood

In this file, it is unclear if “flour” is meant to be distinct from “sweet potato flour”. Perhaps plain “flour” is wheat flour? It is also worth noting that whereas most of the ingredients listed are separate by a comma, here the two flours are separated by the word “and”

### Example:

<https://en.wikipedia.org/wiki/Hominy>

It has a section in the infobox “Ingredients generally used” but not main\_ingredients

Nisholda has both main\_ingredients as well as ingredients generally used:

<https://en.wikipedia.org/wiki/Nisholda>

### Example:

Some common foods don’t have any ingredients listed in the infobox.

For example:

Potato chips: [https://en.wikipedia.org/wiki/Potato\\_chip](https://en.wikipedia.org/wiki/Potato_chip)

(oil would be expected to be listed)

Whereas French fries does have it listed: [https://en.wikipedia.org/wiki/French\\_fries](https://en.wikipedia.org/wiki/French_fries)

### Example:

For Buffalo wings, the ingredients seem to be listed one per line rather than a single line, separated by commas

[https://en.wikipedia.org/wiki/Buffalo\\_wing](https://en.wikipedia.org/wiki/Buffalo_wing)

#### **Main ingredients**

[Chicken wing](#)

[Cayenne pepper](#) hot sauce

[Butter](#)

### Example:

Hot dog ingredients, are listed very “free form”:

[https://en.wikipedia.org/wiki/Hot\\_dog](https://en.wikipedia.org/wiki/Hot_dog)

“Sausage made from pork, beef, chicken, turkey or combinations thereof and a bun”

“bun” isn’t very clear, since it is bread

### Example:

Another free-form entry:

```
| name = Tahu campur  
| main_ingredient = fried [[tofu]] garnished with noodles, bean sprouts, lettuce, and krupuk cracker. Served with  
beef stew in petis-base soup.
```

### Example:

Bread pudding:

[https://en.wikipedia.org/wiki/Bread\\_pudding](https://en.wikipedia.org/wiki/Bread_pudding)

Rather than listing “bread”, it states: “Usually stale bread”

### Example:

When parsing the ingredients, not everything will necessarily be on the same line. For example:

```
| name = Roat  
| image = Dum ka roat.jpg  
| image_size = 250  
| caption = Roat topped with almond and raisins  
| alternate_name = Dum ke roat  
| course = [[confectionery]]  
| served = At room temperature  
| main_ingredient = {{flatlist}}  
*[[Semolina]]  
*[[refined flour|maida (refined flour)]]  
*[[ghee]]  
*[[khoa]]  
{{endflatlist}}  
| variations =  
| calories =  
| other =  
| associated_cuisine = [[Hyderabadi cuisine]]  
| place_of_origin = [[Hyderabad]]  
| minor_ingredient = {{flatlist}}  
*[[Dried fruits|Dry fruits]]
```

```
*[[raisins]]
*[[saffron]]
*[[nutmeg]]
{{endflatlist}}
```

### Example:

This is another example of unusual formatting, although it is all on one line in this case:

```
<text bytes="3242" sha1="krsvcwvt82iadzqw54l5r7pvkixw9dr"
xml:space="preserve">{{Short description|Bhutanese cottage cheese}}
{{Infobox
food|name=Datshi|place_of_origin=[[Bhutan]]|main_ingredient=Milk|image=Datshi.jpg}}
''''Datshi'''' '''([Dzongkha]]:''འཇམ་པལ་'';[[Wylie transliteration|Wylie]]: dar-tshi) is a
traditional [[Bhutan]]ese [[cottage cheese]] commonly produced from cow milk. It is a
naturally fermented milk product produced using the back-
```

### Example:

For the food ‘macaroni’, Chat GPT listed ‘cheese’ as an ingredient, which would have instead been expected for the food ‘macaroni and cheese’. In fact, it listed the exact same set of ingredients for both ‘macaroni’ and ‘macaroni and cheese’.

Chat GPT also listed ‘imitation shark fin noodles’ as an ingredient in ‘imitation shark fin soup’, even though it seems that no such ingredient actually exists.