



HIGHLIGHT

Food information engineering

Azanzi Jiomekong^{1,2} | Allard Oelen² | Soren Auer² | Lorenz Anna-Lena² | Vogt Lars²

¹Department of Computer Science, University of Yaounde I, Yaounde, Cameroon

²TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany

Correspondence

Azanzi Jiomekong, Department of Computer Science, University of Yaounde I, Yaounde, Cameroon.
Email: jiomekong@tib.eu

Funding information

TIB-Leibniz Information Centre for Science and Technology

Abstract

Food information engineering relies on statistical and AI techniques (e.g., symbolic, connectionist, and neurosymbolic AI) for collecting, storing, processing, diffusing, and putting food information in a form exploitable by humans and machines. Food information is collected manually and automatically. Once collected, food information is organized using tabular data representation schema, symbolic, connectionist or neurosymbolic AI techniques. Once collected, processed, and stored, food information is diffused to different stakeholders using appropriate formats. Even if neurosymbolic AI has shown promising results in many domains, we found that this approach is rarely used in the domain of food information engineering. This paper aims to serve as a good reference for food information engineering researchers. Unlike existing reviews on the subject, we cover all the aspects of food information engineering and we linked the paper to online resources built using Open Research Knowledge Graph. These resources are composed of templates, comparison tables of research contributions and smart reviews. All these resources are organized in the “Food Information Engineering” observatory and will be continually updated with new research contributions.

INTRODUCTION

According to WHO, every country in the world is affected by one or more forms of malnutrition (WHO 2023). In fact, inadequate food/nutrient intake results in three forms of malnutrition: (1) undernutrition including wasting (low weight-for-height), stunting (low height-for-age) and underweight (low weight-for-age), (2) micronutrient-related malnutrition: this includes micronutrient deficiencies (a lack of important vitamins and minerals) or micronutrient excess, (3) Overnutrition including overweight, obesity, causing diet-related noncommunicable diseases such as heart disease, stroke, diabetes, and some cancers (WHO 2023). Globally, overweight and obesity

have been found to be one of the major risk factors for various chronic diseases such as diabetes and cardiovascular diseases (WHO 2023). In developing countries, malnutrition remains a major challenge to public health where the childhood malnutrition rate often reached 35% (Bitá et al. 2020) and the risk factors associated with children includes Malaria, cholera, and other diseases. Concerning adults, Africa is experiencing a rapid increase in overnutrition linked to the development of noncommunicable diseases while the main diet-related problems involve obesity, cardiovascular diseases, diabetes, and hypertension (Owino 2019). However, adequate nutrition is an essential catalyst for economic and human development as well as for achieving Sustainable Development Goals.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *AI Magazine* published by John Wiley & Sons Ltd on behalf of Association for the Advancement of Artificial Intelligence.

Understanding Food information can allow people to have a healthy diet and avoid malnutrition. To this end, food information engineering involves the acquisition, the processing and the diffusion of up-to-date food information to different stakeholders (Jiomekong 2023b). This information is compiled from several data sources and used for a variety of purposes such as food recommendation (Alert et al. 2022; Brintha et al. 2022; Chen et al. 2021), ingredient substitution (Min et al. 2022), food image recognition (Anthimopoulos et al. 2014; Bettadapura et al. 2015), and so forth. Many authors have proposed statistical and AI technologies for food information engineering. AI technologies involve symbolic AI, connectionist AI and neurosymbolic (NeSy) AI (Goel 2022). Symbolic AI posits the use of background knowledge (in the form of symbols in relation—also called concepts in relation) for generating new knowledge through inference and reasoning. Connectionist AI relies on learning associations from data. Unlike symbolic AI, connectionist AI needs large-scale data to build AI models. These models consider that knowledge is distributed (rather than being localized) and retrieved through spreading activation among connections. Neurosymbolic AI consists of the combination of connectionist models with symbolic models in a way that retains the strengths of each paradigm while overcoming their complementary weaknesses (Goel 2022). This gives rise to many advantages (Pan et al. 2023) such as the ability to learn with less data, robustness to noise, the ability to perform joint reasoning (structured prediction), and so forth. This is very important for food information engineering because of the multimodal form of food information.

A huge number of research papers have been published in the domain of food information engineering, each paper covering different aspects. This huge number of publications poses significant challenges for the researchers who want to sort through, read, understand, compare, and build upon to determine for instance, the state-of-art of the domain. The problem of access to scientific knowledge disseminated in scientific repositories is a well known problem (Auer et al. 2020; Nasar, Jaffry, and Malik 2018; Shah et al. 2003; Jiomekong, Auer, and Oelen 2024) in the research community. To solve this problem, state-of-the-art research is generally published (e.g., Min et al. 2022, 2019, etc.) However, these papers cover only some aspects of the domain. On the other hand, these reviews are static resources which are updated only by creating new versions. The creation of new versions may take time and the review becomes obsolete. This paper aims to cover the different dimensions of food information engineering. The main innovation this paper brings in the writing of research surveys in general is the linking of the survey (which is a static resource) to dynamic resources

which are progressively updated by the researchers of the domain, forming the linked open literature review also called semantic review or review 3.0 (Jiomekong, Auer, and Oelen 2024).

To allow the food information engineering researchers to collaboratively build the body of knowledge of the domain, we propose to extract and organize the research contribution of authors in a scholarly Knowledge Graph (KG). Thereafter, these research contributions can be compared using annotated tables and graphics. The latter are used to write smart reviews. We particularly use Open Research Knowledge Graph¹ (ORKG) (Auer et al. 2020). To organize the food information engineering research contributions using ORKG, we propose a set of resources composed of ORKG templates, comparison tables and smart reviews. All these resources are organized in the ORKG observatory named “Food Information Engineering” observatory. This observatory contains dynamic resources that will be continually updated by the researchers of the domain.

In the rest of this paper, we firstly present an overview of ORKG and the research methodology. Thereafter, we present the research results organized into collecting food data, organizing food data, processing food data, and using food data. Finally, future research direction, the discussion and conclusion are presented.

OPEN RESEARCH KNOWLEDGE GRAPH

The ORKG is a *neurosymbolic knowledge graph* designed to semantically structure and interconnect scholarly information, blending the strengths of symbolic AI with neural network approaches (Auer et al. 2020; Jiomekong, Auer, and Oelen 2024). It is built according to the principles of Open Science, Open Data, and Open Source. Thus, all the resources are freely available for update and use and reuse (Stocker et al. 2023). To date, ORKG comprehensively describes more than 25,000 research papers addressing more than 6000 research problems (in 700 research fields), resulting in more than 1500 state-of-the-art comparisons, 400 templates, contributed by 1600 users, and 80 smart living reviews.² In the following paragraphs, we present the main ORKG features being used for food information engineering literature review.

To document a domain, the first step consists of adding papers in the system³ and providing the semantic description of these papers. In this work, we recommend the use of templates. The latter specifies the structure of scientific information so that fellow researchers can complete with more key-insights and new researchers can rapidly get insights into the research domain. Figure 1 presents an example of a template.⁴ The left side contains

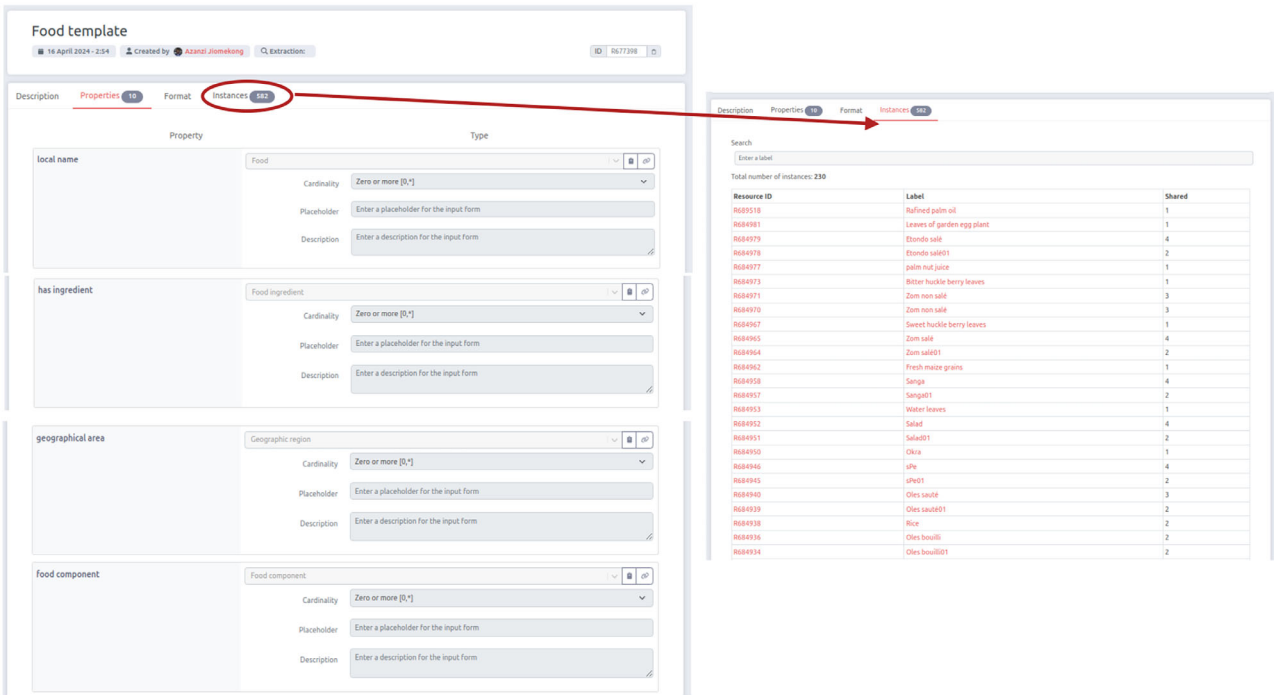


FIGURE 1 A template for describing foods (on the left) and food instances (on the right).

Properties	Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions <i>Cal estim = cal(AmericanDS - Multi-task CNN) - 2017</i>	Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions <i>Cal estim = cal(JapaneseDS - Multi-task CNN) - 2017</i>	Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions <i>Cal estim = cal(AmericanDS - Multi-task CNN) - 2017</i>	Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions <i>Cal estim = cal(JapaneseDS - Multi-task CNN) - 2017</i>	Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions <i>Cal estim = cal(AmericanDS - Multi-task CNN) - 2017</i>
training data size	70	70	70	70	70
absolute error (kcal)	128.5	95.2	120.5	94.4	119.6
correlation	0.293	0.802	0.361	0.811	0.383
dataset	American food DS	Japanese food DS	American food DS	Japanese food DS	American food DS
evaluation data size	30	30	30	30	30
machine learning method	Multi-task CNN	Multi-task CNN	Multi-task CNN	Multi-task CNN	Multi-task CNN
relative error (%)	43.3	27.9	42.9	27.6	41.7
research problem	Food calorie estimation	Food calorie estimation from food image	Food calorie estimation from food image	Food calorie estimation from food image	Food calorie estimation from food image

FIGURE 2 Comparison of machine learning models used to estimate calorie from food image.

properties for comparing food and on the right, research contributions created using this template.

Knowledge extracted and organized can be used to build comparison tables⁵ and write smart reviews.⁶ Figure 2 presents the comparison of machine learning models used to estimate calories from food image. Figure 3 presents an example of a review.

Through the ORKG observatory, a whole community of researchers can gather and build the state-of-the-art of their domain. This observatory contains research papers, comparison tables and reviews of the domain. They can be added to literature reviews and published in conferences and journals. If new literature is published, it is easy

to continuously expand the comparison tables and smart reviews, which thus continues to reflect the state-of-the-art of the domain. Figure 4 presents the food information engineering observatory.

RESEARCH METHODOLOGY

The research methodology (Jiomekong and Tiwari 2024) (see Figure 5) consists of an iterative and incremental process during which knowledge are extracted from scientific papers, organized using ORKG and published. It was divided into two phases: an exploratory phase and a

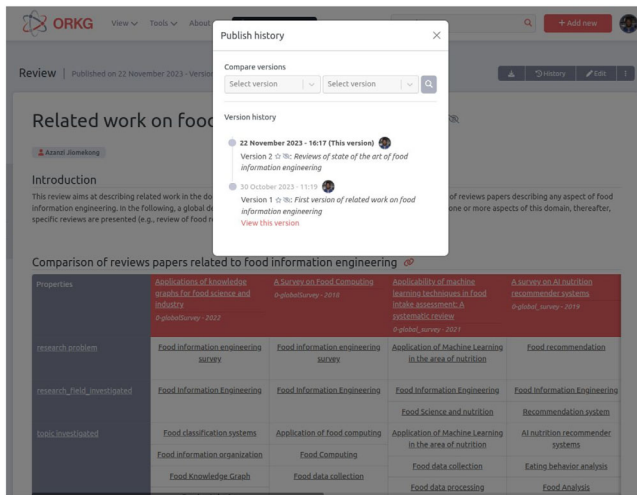


FIGURE 3 An example of a review.

systematic literature review phase. The exploratory phase took place over 6 months (June–November 2022) curation of ORKG with the help of the ORKG team. The curation meetings allowed us to benefit from the remarks of the other curators and the ORKG team. The overall work was evaluated during this exploratory phase. The systematic literature review phase consists of using the resources created during the exploratory phase to document more papers and create new resources and continuously update existing ones. Figure 6 presents an overview of the resources created during this work.

Globally, the research methodology consists of six steps during which classes, properties and relations are extracted from scientific papers and used to build templates. Thereafter, these templates are used to represent contributions of papers related to the same research problem. Finally, the contributions are used to build comparison tables, which themselves can be used to write smart reviews. The following paragraphs give an overview of each step.

Knowledge elicitation

Knowledge elicitation is the most laborious task of this work. It consists of identifying the research domain to document, research problems that are involved in this research domain and use these information to search for scientific papers using search engines and digital research repositories. In this work, for each research problem (e.g., “food recommendation”, “food image recognition”, “food ontology construction”), several papers were identified, read, keywords, clauses, sentences, scientific claims, globally, all the information that is relevant to understand

the paper were extracted. Sometimes, knowledge identified can refer to entities which are grouped to give classes. These classes are put in relation to each other. At the end of this step all the relevant knowledge is extracted. Globally, around 230 papers were identified and used.

Knowledge analysis and interpretation

Knowledge analysis and interpretation consists of reviewing the elements extracted, deleting irrelevant and duplicated terms and providing a definition of each term. Knowledge obtained after this activity is classified into classes, relations, properties and instances. Around 24 new classes (e.g., “Type of food⁷”, “Food component⁸”) and 32 new properties (e.g., “list food components⁹”, “food group¹⁰”) were created and used. Around 22 and 24 existing classes and properties, respectively, already created in the ORKG system were reused.

Template creation/update

The classes, properties, relations and instances are used to create templates using the ORKG template editor. These templates are conceptual models of papers dealing with the research domain and research problem addressed by its creator. It is used to semantically describe the papers by following the Subject-Predicate-Object triple principle. During this step, existing templates can also be updated with new insights. For instance, instead of creating a new template for documenting food ontologies, we updated an existing one. In total, seven templates were created and four already created were reused.

Knowledge representation

The knowledge representation step consists of using ORKG templates to annotate research papers related to the research domain and the research problem. We consider in this research that all the key insights in the research paper such as the definition of research problems, materials and methods used, results obtained, lessons learned, and so forth are grouped into research contributions. These research contributions are created using related templates. For instance, the research contribution describing food image dataset Food 101 is described using the template “food image dataset¹¹”. More than 467 research contributions from 230 papers, related to around 16 research problems were created.

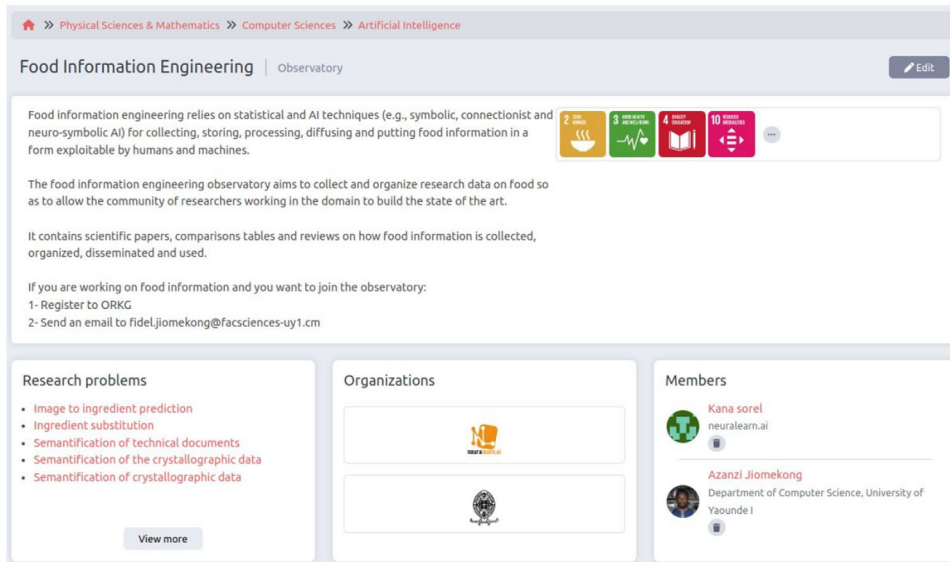


FIGURE 4 Food information engineering observatory.

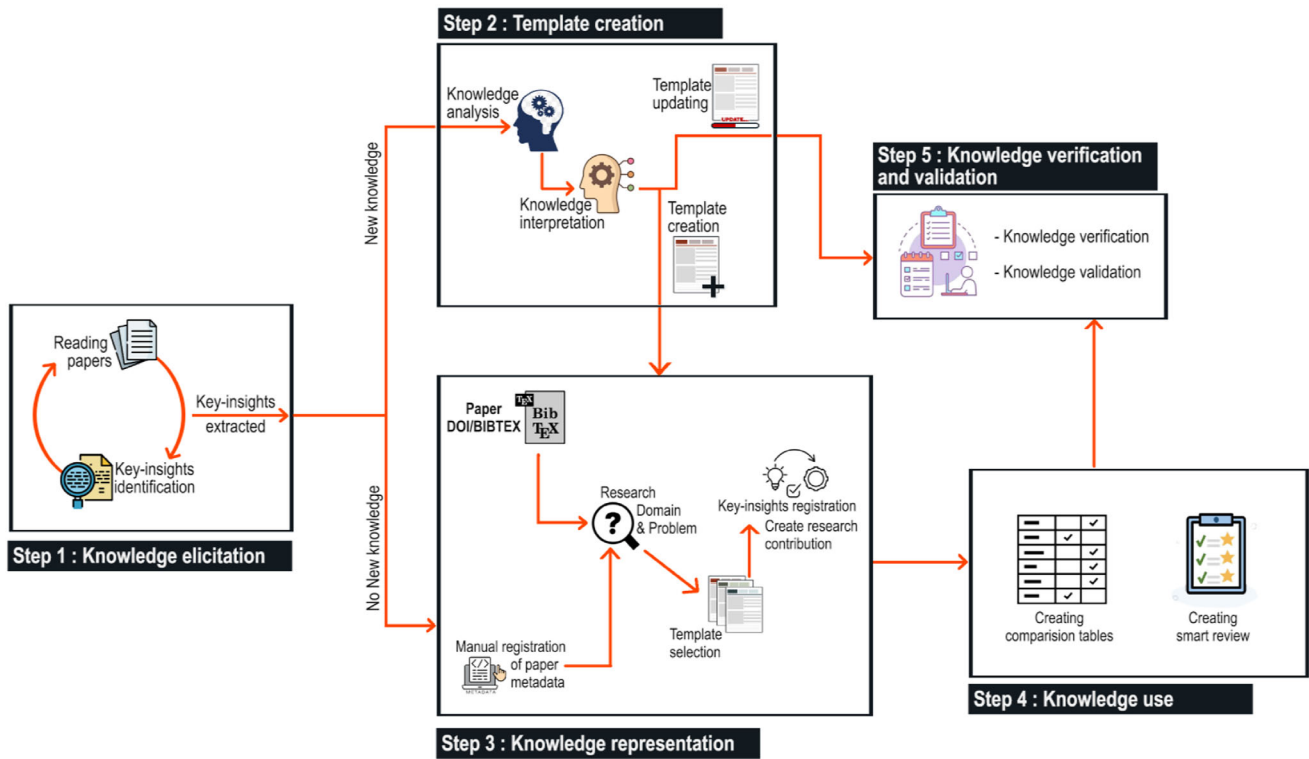


FIGURE 5 Research methodology.

Knowledge use

The structured content descriptions of scientific contributions are presented in such a way that the research contributions (from the same or different research papers) become comparable with others. Thus, these research contributions are used to build comparison tables using ORKG

comparison tables wizard. After the creation of several comparison tables, a researcher may create smart reviews for giving an overview of research addressing a particular research question. For this task, ORKG furnishes a “What You See Is What You Get (WYSIWIG)” editor. In this work, around 40 comparison tables and 9 smart reviews are created.

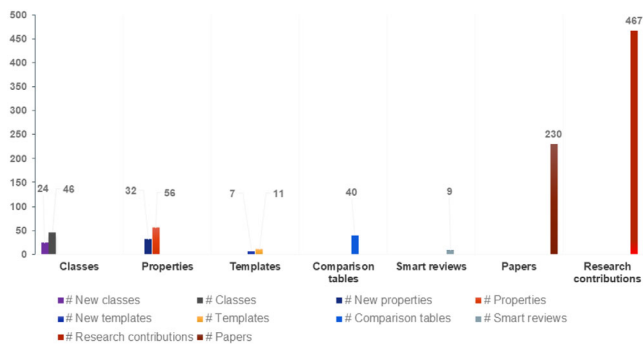


FIGURE 6 Overview of ORKG resources created or reused.

Collaborative work on literature review

To allow the food information engineering community to collaborate to organize the domain and ensure height quality standards, the food information engineering observatory was created.¹² Scientific papers, Templates, research contributions, comparisons tables and smart reviews are organized in this observatory. These resources constitute dynamic resources to which this paper is linked to. Thus, researchers can gather these resources to make them more representative of the food information engineering domain. The template can help newcomers to know what are the key -insights that are searched by fellow researchers and to experienced researchers to improve with more research contributions to the document. If new research papers are published, anyone can add new contributions to comparisons tables, publish new versions of these tables and update the smart reviews.

COLLECTING FOOD INFORMATION

Data is at the center of food-based AI systems. These data come from food data sources. Food information are extracted from these data sources and used to build food datasets. Thus, the first result of this work consists of presenting how food data is collected from data sources and used to build food datasets. To keep track of the evolution of this result, a smart review is provided as reference for more details and also, to have up-to-date information (Jiomekong 2023a).

Food data sources

Thanks to the deployment of the Internet, various smart devices, Internet of Things (IoT), and networks such as social networks, mobiles networks, a great amount of food data is being recorded in various modalities

such as text, images, videos, and sound. In this section, we review the different data sources that are generally used for food data collection and briefly give the summary. We organized these sources into the following dimensions:

- Human sources:** Humans are the principal source of food data. The majority of food data sources are created and updated by humans. Thus, humans can be a valuable source of information. In many cases (e.g., food ontology construction Jiomekong 2022c), information are directly acquired from domain experts (e.g., experts in food science and nutrition). In the domain of food science and nutrition, food components are generally obtained through chemical analysis of food (Greenfield and Southgate 2003). Domain experts may also play the role of data validator to ensure that data collected is related to the research domain (Jiomekong et al. 2023). In addition, in many cases, food data are recorded via food logs (Metwally et al. 2021). The latter records what one eats or drinks daily.
- Structured sources:** Structured sources such as databases, Knowledge Graphs can be used as relevant food data sources. Official organizations (e.g., FAO) and many countries provide food composition tables in tabular format. In addition, food vocabularies, ontologies and KG are useful food data sources. Some examples involve USDA, Odor threshold database, food product codes, and so forth.
- Semistructured sources:** These sources provide a structure that can be exploited to design systems for automatic extraction. For instance, food descriptions are stored in the form of tables in scientific papers (Jiomekong et al. 2023) and food websites (Yummly,¹³ Tripadvisor,¹⁴ etc.) which provide food information following a structure (HTML). These structures can be exploited to build automatic tools for automatic food information extraction.
- Unstructured sources:** Unstructured sources do not provide information with any structure that can be identified at first glance. Unstructured food sources concern data sources containing free text (for instance, cooking recipes), food images and food videos. This data is being published daily by consumers on the Internet.

Food data acquisition

Data can be acquired manually, automatically or semiautomatically (Jiomekong, Camara, and Tchuenta 2019) and the food data acquisition is no exception. In the following points, we present data acquisition by considering the three dimensions of the data sources.



- **Human sources:** The acquisition of food data from human sources is always manual because people from which information is coming from should provide this information by observation, talking or writing. In the early years, food data were acquired from human experts or from people using questionnaires. However, with the widespread use of smart devices, IoT toolkits, social networks, and so forth many other ways of acquiring food data emerge. For instance, a user can use a smartphone to register what he is eating daily and build his food log (Metwally et al. 2021). In certain cases, crowdsourcing platforms such as Amazon Mechanical Turk (AMT) is used for collecting food data (Min et al. 2019). It should be noted that data acquisition through human sources is time-consuming, laborious and hard to achieve at large-scale.
- **Structured sources:** The organization of data from this type of data sources (CSV, JSON, XML, etc.) make it easy to build automation tools for data acquisition. For instance, to build food ontologies, one may rely on structured sources such as databases and use automatic tools to automatically extract food knowledge (Jiomekong 2022c).
- **Semistructured sources:** Much food information is embedded in web pages. The latter follows a structure (HTML) that makes it easy to build automatic tools (based on web scraping) and extract food information easily. The table structure of food composition tables stored in scientific papers make it easy to build automatic OCR tools for its extraction (Jiomekong et al. 2023).
- **Unstructured sources:** Information extraction from unstructured sources such as free text, images, and videos is the most difficult and challenging due to the nature of these information. For instance, food images are different from the other types of images (Min et al. 2019). Many of them do not have a distinctive spatial layout, they have deformable food appearance and thus lack rigid structures. In addition, ingredients are generally randomly distributed in dishes.

Generally, one distinguishes manual acquisition and automatic acquisition (Jiomekong 2023a). Manual acquisition can be used for instance, to annotate food images, by a human who identifies the food and labels the visible food ingredients. However, to build large-scale datasets, this method is very difficult. Fortunately, the development of connectionist models, based on Neural Network, provides useful support for food data acquisition from unstructured sources.

During the data acquisition activity, the data collector should be careful of several types of problems and challenges that may arise. For instance, food composi-

tion tables may have different table layouts, some data may be misspelled, duplicated or there may be ambiguity in some terms. Some food images may be corrupted or damaged, food images do not exhibit any distinctive spatial layout and ingredients from many types of food images are generally distributed randomly on a plate. Thus, the data collectors should define a clear data collection protocol that they will follow during the whole process.

Food datasets

Once acquired, food information are organized into datasets for different purposes. Given the multimodal nature of food information, we consider two types of food datasets: unimodal datasets and multimodal datasets (Jiomekong 2023a).

- **Unimodal datasets:** Contains data of only one type. For instance, many recipe datasets (e.g., Recipe 1M, Recipe Question Answering Min et al. 2022), food composition tables datasets (e.g., TSOTSATable Jiomekong et al. 2023), odor threshold datasets are composed of data in text format (Min et al. 2022). Many food image datasets (e.g., UEC Food100) are composed of sets of images and many cooking videos datasets (e.g., EPIC-KITCHENS) are composed of video data (Min et al. 2022).
- **Multimodal datasets:** Multimodal datasets contain structured data such as symbolic representation of some food in form of knowledge graph, ontology, and so forth and image (or videos) of the corresponding food. The symbolic representation contains basic information such as the list of ingredients, the recipe associated, the cuisine category, the course, flavor, food composition.

FOOD INFORMATION ORGANIZATION

The datasets obtained after the collection can be seen as information silos because they are a collection of data that is not easily or fully accessible. This limits the ability to exploit food data in its full potential. Food information organization aims to put the food information in a structured way and ideally, machine readable. Food information can be organized using tabular data (e.g., food composition tables or databases and many food logs) or using AI techniques. AI techniques for organizing food information consist of symbolic organization, connectionist organization and neurosymbolic organization. An online review to keep track of the evolution of food information organization is provided as reference (Jiomekong 2023c).

TABLE 1 An example of a Food Composition Table of some Cameroonian foods.

	Dietary reference intake	Ekomba	Ekwang	Tenue militaire	Koki
Protein	13 g/day	38.2	10.8	16.4	41.6
Fat	30 g/day	26.6	14.2	17.5	19.5
Dietary fiber	19 g/day	9	1.1	4.1	4.9
Carbohydrates	130 g/day	17.1	11.9	11.1	11.7
Energy	1350 cal/day	13.4	7.8	8.4	10
Calcium	500 mg/day	6.1	2.7	4.9	7.8
Magnesium	80 mg/day	38.4	16.2	21.6	30.6
Sodium	1000 mg/day	56.8	33.6	37.5	51.6
Potassium	3000 mg/day	3.2	4.2	2.1	6.1
Iron	7 mg/day	7.2	16.6	21.6	19.4
Zinc	3 mg/day	21.2	8.9	10.6	37.8
Copper	0.34 mg/day	36.2	55.9	49.9	36.7
Manganese	1.2 mg/day	29.5	23.3	34.4	31.5
Vitamin A	0.3 mg RAE/day	35.9	108.8	119.2	119.1

Tabular organization

In many cases, food information is organized in a tabular form. For instance, many food-related software such as FoodLog Apps (Metwally et al. 2021) are used to follow the nutrition of people in diet. They store food information using a relational database. Thereafter, they use SQL (or equivalent) language to query the database for several purposes such as the quantity of calories consumed by a user, to calculate the user body mass index (BMI), etc. To solve the problem of malnutrition, several efforts have been made worldwide. One consists of compiling data about food, food ingredient, food component into usable databases such as Food Composition Table (FCT) or Food Composition Databases (FCDB). Table 1 presents an example of a FCT of some Cameroonian foods. It shows for instance that the consumption of Ekwang, Tenue militaire and Koki (100 g) by children aged between 1 and 2 years would meet more than 100% of their daily recommended intake for vitamin A. These foods can thus be recommended for children suffering from vitamins A deficiency (Ponka et al. 2016).

FCT are built at the countries or regional levels. Ideally, each country should dispose of a Food Composition Table, even if this is not always the case. The Food and Agriculture Organization (FAO) of the United Nations keeps track of existing FCT (FAO 2022). To help countries develop their FCT, FAO provides FAO/INFOODS compilation tools in MS Excel format for storing, documenting and managing FCT electronically. This standard has been used to build several food composition tables. But in many cases, countries developed their own FCT and stored it in relational databases (FAO 2022).

Even if FCT allows nutritionists dispose of a useful tool to assess the food one eats, many of them suffers many problems such as (1) static databases sometimes in print form or in Excel format; (2) many developing countries don't have reliable FCD and the majority are outdated; (3) FCT is not easily accessible to the general public. However, if well presented, the general population can use the system for assessing nutrient intake, get access to dietary advice, prevent diseases, and so forth.

Symbolic organization

To organize food information, symbolic organization uses symbols to represent background food knowledge. Due to the large-scale nature of these data, they should be organized so as to be in a machine and human-readable form and facilitate their exploration and exploitation. To this end, food information are linked together forming either food classification systems, food ontologies, food knowledge graphs or food linked data (Min et al. 2022). An example of symbolic representation of some Cameroonian foods and their composition is given in Figure 7.

Food classification and description systems (Jiomekong 2023c) organize food knowledge in a hierarchical form, providing basic standards for organizing and communicating food information. They are generally constructed for a well-defined purpose such as dietary monitoring, nutrient risk assessment, exposure assessment.

Food ontologies (Jiomekong 2022c) are used to describe food concepts, their relations, axioms and rules. They are used to provide a shared understanding of food knowledge, to tackle data harmonization problems that may

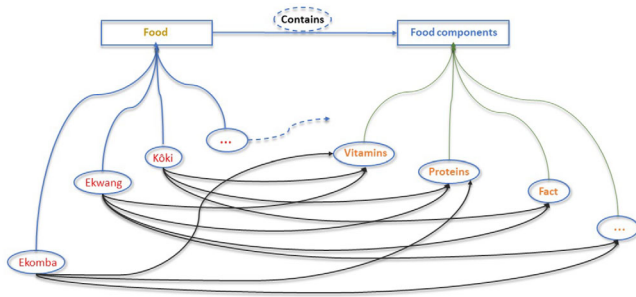


FIGURE 7 An example of symbolic organization of some Cameroonian foods.

arise between different food related domains such as nutrition, agriculture, etc. Several food ontologies have been developed and published for different purposes such as knowledge graph construction, nutrition monitoring, food traceability, and so forth.

Food KG (Jiomekong 2022b) describes real-world food entities and their relations. In many cases, food knowledge graphs use ontologies as their schema. In these cases, the related ontology is populated with instances to form the food knowledge graph. Food KGs can be classified into unimodal and multimodal KGs (Min et al. 2022, 2019).

- **Unimodal food KGs:** These knowledge graphs are mainly verbal knowledge graphs and contain entities in the form of text connected together.
- **Multimodal food KGs:** To support the multimodal nature of food information, a multimodal food knowledge graph is constructed. Such a knowledge graph may incorporate visual (images), textual, structured data, rich context information, as well as their diverse relations.

Food-oriented linked data aims to transform food information silos to a more reusable globally digitally connected Internet of Food (Min et al. 2022). To this end, food concepts are linked to existing knowledge graphs such as DBpedia, Wikidata or ontologies such as FoodOn. The TSOTSATable dataset, containing food and their compositions is linked to Wikidata, ORKG KGs and FoodOn ontology (Jiomekong et al. 2023).

Connectionist organization

Connectionist organization of food information consists of learning associations from food data and storing this information in the form of connections between neurons. To this end, it requires a lot of data. Using these data, it adjusts the strength of the connections (weights) between its nodes (or neurons). The main problem currently in

the domain is the lack of food data (Min et al. 2022, 2019). In this case, existing connectionist models trained on similar data are fine-tuned. Currently, several connectionist models are proposed for food recognition (such as VGG-19, AlexNet, GoogLeNet, Resnet-50, DenseNet, MobileNets, ShuffleNets), food recommendation, recipe prediction (Jiomekong 2023c), and so on.

Neuro-symbolic organization

NeSy proposes the combination of connectionist and symbolic organization and simultaneously leverages their advantages. Concerning food information engineering, this can be done for many reasons:

- **Multimodal nature of food information:** Food information involves food images and text. Currently, connectionist models are the state-of-the-art for image recognition. Thus, one may combine a connectionist model for food recognition with the symbolic model (e.g., a food composition knowledge graph) for nutrient prediction.
- **Explainability:** Because the processing of connectionist models is often opaque, infusing symbolic knowledge into these models may help to understand the predictions obtained. One may combine a Large Language Model (LLM) for food Question Answering with a food KG in order to understand the prediction of the LLM.
- **Enable multistep inference:** The connectionist model can be trained on symbolic models such as KG to enable multistep inference. An example of a three-step inference: from a food name/image, infer the list of ingredients composition; from the food ingredient, infer the nutrient; from the nutrient, infer if the food can be recommended to a person in diet.

Although several works have been done on representing food information using symbolic and connectionist AI, neurosymbolic AI is still rarely used. The Figure 8 presents an illustration of food information that can be organized in a Neurosymbolic Knowledge Graph (NeSyKG). This graph is composed of the following models:

- A connectionist model for images representation,
- A Large Language Model containing huge general knowledge,
- A Knowledge Graph containing explicit knowledge.

The relations in these models show how they work in synergy to respond to the user's queries. For instance, the connectionist model representing images may be used to recognize the food, the LLM used to

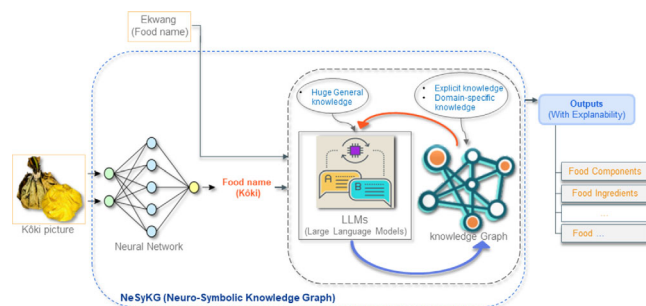


FIGURE 8 Illustration of the organizing food information in a Neuro-symbolic Knowledge Graph.

determine food composition and the KG provides explainability.

FOOD INFORMATION PROCESSING

Information/data processing can intervene at any step of the food information engineering workflow. Data processing can be done manually, semi-automatically or totally automatically. In this paper for instance, we are using manual processing of information extracted from scientific papers to enrich the ORKG knowledge graph. The manual processing of information is tedious and difficult to apply at large scale. This section focuses on automatic processing of food data. It considers statistical methods and AI methods. AI methods rely on symbolic, connectionist and neurosymbolic AI. To keep track of the evolution of the food data processing, we provide as reference a smart review that will continue to be updated (Jiomekong 2022a).

Statistical methods

Statistical analysis is the main tool used by researchers in food science to analyze the data collected in the fields and answer the research question. Statistical methods have been widely applied to enhance several food-related tasks (Jiomekong 2022a). Multivariate data analysis is used to predict reasonable sensory properties from chemical characterization of sauces. The combination of Bag of Features (BoF) and Support Vector Machine (SVM) is used to distinguish food images from other images, and retrieve personal food images for food balance estimation. Bag of Textons is combined with Scale-Invariant Feature Transform (SIFT) for food image retrieval by comparing the food images through similarity measures. Statistical method employed co-occurrence relation information between food items for recognizing multiple-food meal photos.

Symbolic methods

Symbolic methods process symbols, which represent real-world concepts and their relationships (Goel 2022). They use logic-based programming where rules and axioms are used to make inferences and deductions. Symbolic methods are highly interpretable, as one can easily trace the reasoning process back to the logical rules that were applied. However, it requires complete and well-defined knowledge to function correctly. In addition, it can become computationally expensive as the number of symbols and rules increases, making it difficult to scale to large the whole symbols of the domain (Pan et al. 2023).

Concerning food information engineering, inference engines are used to generate new facts from symbolic representations of knowledge such as food ontologies and food knowledge graphs. Using existing facts, symbolic methods use semantic-based inference to generate new facts and make predictions. To this end, inference engines such as ELK, FaCT++ Pellet, and so forth may be used. Symbolic methods have been used to support complex food query, food Question Answering and multimodal food dialog via inference engines (Min et al. 2019).

Connectionist methods

Connectionist methods process data (or are trained on data) to build connectionist models representing these data. This consists of learning associations using food data and storing this information in the form of connections between neurons. These methods consider three elements, the input data (data to be processed), a connectionist model (model used to represent these data) and the output (the resulting connectionist model). The system takes as input the data which can be text, images, and others; process this data and return a model representing the data. The resulting model can be used for several tasks. Connectionist models (CNN, GoogLeNet, Resnet-50, AlexNet network, etc.) have proven their superiority in several task (Jiomekong 2022a) such as food recognition, ingredient detection, food segmentation, food volume estimation, food recommendation, food calorie estimation from food image.

Connectionist methods are also used in AI applications such as mobile applications (Jiomekong 2022a) to process information before being returned to users. In fact, with the widespread use of mobile applications, many connectionist models are trained and used for food information processing and prediction. The application of these systems includes food recognition, restaurant-specific food recognition.



In many cases, data may be insufficient for training a model by zero. In this case, existing models are fine tuned. YOLOv2 has been fine-tuned for food detection with multiple food items (Min et al. 2019).

NeSy methods

Neurosymbolic methods rely on the combination of connectionist methods and symbolic methods for food information processing. For instance, to infer ingredient and/or food composition from a dish image, a deep neural network can take as input the food image and return as output the food name. Thereafter, the food name can be used as input to a knowledge based system that uses an inference mechanism to infer the food ingredients and food components. In many cases, connectionist models learn feature representation of KGs and identify errors, infer new conclusions, enrich the food KG with new relations predicted. For instance, joint deep feature learning from GoogLeNet and semantic label inference have been used for food recognition; bipartite-graph labels have been combined with CNN in a unified framework for multilabel ingredient recognition (Jiomekong 2022a).

USING OF FOOD INFORMATION

Food information processing allows us to put information in a form that can be used by different stakeholders for a variety of purposes such as food search, food recommendation, food Question Answering, decision making, etc. Depending on the user of the food information, the latter will be put in an appropriate format and diffused using appropriate channels. In this section, we present the different users of food information, how food information is generally diffused and the different use cases. A smart review allowing to keep track of research on the use of food information is provided in reference Jiomekong (2022d).

Stakeholders

All the people in the world are involved in the production, processing, and use of food information. Given different usage, we classified them into the following categories:

- **General population:** This category involves all the people in the world. The food information is generally used by these people to choose their food given to food perception, taste, preferences, or their health status.
- **Health professionals:** Health professionals generally use food information for identifying the origin of a

health problem such as allergy, foodborne diseases, etc. To trace back and understand the origins of certain symptoms, health professionals generally ask questions on the eating history of the patients.

- **Nutritionists:** This category of stakeholders uses food information for nutrition advice. Many people suffer health problems that need nutrition monitoring such as diabetes, overweight, cardiovascular diseases. Nutritionists are specialists who generally follow the diet of these people by using tools such as food logs.
- **Decision makers:** Decision makers make use of food information to ensure that the population has safe and enough food. For instance, information on food production may allow decision makers to make decisions and avoid lack of food.
- **Food manufacturers, distributors, and retailers:** Knowledge on the eating behavior of the population can help this category of users to identify a geographical area in which period which kind of food can be proposed to the customers. In addition, a better understanding of the process used by people to assess the acceptability and flavor of new food products may be used by food manufacturers to produce acceptable food.
- **Researchers:** Food information engineering is a multidisciplinary research domain in which many types of researchers are found. It involves researchers from food science and nutrition, food chemistry, microbiology, computer science, agriculture. These users make use of food information to draw and/or validate hypotheses, build AI models, make predictions.

Diffusion of food information

Effective communication strategies must address the needs of a heterogeneous group of users of food information. To this end, food information should be diffused to different stakeholders according to their profile. For each profile, appropriate presentation of information (food data visualization) and channel should be used. Statistical tables, graphics and pictures are used for food information visualization to decision makers and by the scientific community.

Concerning the general population, information is shared using many channels such as social media, television, radio. For instance, in many television channels, many diet programs recommend eating fruits and vegetables every day. Many populations in the world, particularly in developing countries or countries in crisis, do not always have access to mass media. However, with the widespread use of mobile networks, SMS messages are used for health information diffusion. We believe that this method can also be used for food information diffusion. The recent

development of Knowledge Graphs provides new ways to present information on the web to a heterogeneous group of users using so-called “knowledge cards” (Min et al. 2022). Combined with visual graphs, knowledge cards allow users to explore food information comprehensively, by clicking and jumping from information to information.

Use cases

The food information is used for several purposes such as food search, food recommendation, food Question Answering. This list is not exhaustive. The following paragraphs present some use cases and the overall use cases will be continuously documented using an ORKG smart review (Jiomekong 2022d).

Agriculture

In the domain of agriculture, food information is collected and processed to improve productivity and quality of crops. For instance, to improve the quality of crops, precision agriculture uses food and geographical information to increase agricultural productivity while reducing adverse impacts on the environment (Marinello et al. 2023; Alazai et al. 2024). Plantation and orchards with dense fruits trees may be difficult to manage. Thus, large datasets of fruit images are processed using machine learning techniques for automatic fruit classification, fruit counting, plant disease detection in large farms (Zhang et al. 2022; M.Pyngkodi et al. 2022; Liu and Wang 2021; Bakr et al. 2022).

Food search

Food search involves single-modality based search and cross-modal based search (Min et al. 2022). Single modality based search (e.g., visual food search or recipe search) consists of searching one type of information (e.g., food image or food recipe). Cross-modal based search involves two types of food information. For instance, searching food recipes/ingredients using a food image, searching restaurant locations using food images.

Given the huge quantity of food information that is published on the Internet daily, the different stakeholders may spend much time searching for food information. In the general population (who generally lack professional knowledge of healthy diet), an overweight person may be searching for dietary information. They may be looking for what kind of food to choose or avoid when they feel

uncomfortable. The semantic food search solves this problem by providing in addition to keywords-based search, the possibility to determine the intent and contextual meaning of the query words. Thereafter, to provide more structured search results and better summaries in the form of knowledge card (Min et al. 2022). In many cases, user profiles, preferences and context information are used to provide more accurate food information to users (Zulaika, Gutiérrez-Fandiño, and de Ipiña 2018).

Food recommendation

Food recommendation systems (Chen et al. 2021; Min et al. 2022) aim to recommend food to users according to several criteria/constraints such as food preference, ingredient availability, nutritional parameters (macronutrients and micronutrient deficiencies or excess), temporal and spatial context, culture, etc. To recommend healthy food, some of them incorporate nutritional aspects (a so-called “calorie balance function”) that take into account the fact that users may have many constrained needs such as allergies or lifestyle preferences, the desire to eat only fruit or vegetarian food, maybe suffering diseases such as overweight or diabetes, or may have a particular profile such as weightlifting athletes, older adults. Food recommendations can go further than simple recommendations of healthy food, but also the restaurant recommendation. For instance, recommending restaurants serving vegetables (Alert et al. 2022). To help diabetes patients to maintain a healthier diet, several authors propose to automatically track the number of food calories they consume daily, along with their total calories (Brintha et al. 2022).

Food question answering

Food Question Answering (Food QA) systems aim to search for food information using queries in natural language (Chen et al. 2021). It handles different categories of questions about recipes and nutrition such as simple queries for nutritional information, comparisons of nutrients from different foods, and constraint-based queries to find recipes matching certain criteria. For instance, a diabetic patient may ask the question: “what is healthier food for a diabetic?”. The system will use several information such as geographic position of the person, culture, food preferences to recommend healthier food to this person. In addition, the Food QA system may propose ingredient substitution to users given food preferences or the health profile of the latter.



Clinical practice

During clinical practices, the consumption of food may be used to understand the cause of health problems. Knowledge about food allergies or non communicable diseases can be used to understand certain symptoms and provide a diet plan to the patient. The health professional can use AI technologies for predicting and monitoring people following certain diet plans. For instance, in food allergies, information on the quantity of food consumed, in which form the food was consumed and the health profile of the person who consumed the food are generally considered (D'Auria, Pendezza, and Zuccotti 2020). Evidence on the role of cows' milk in children with constipation is used by clinicians to identify constipation problems from children (Connor et al. 2022).

Food recognition

Nowadays, people travel from one place to another more easily than ever. However, food consumed in different geographical areas is not always the same. It can thus be interesting to people when moving from one place to another to have a better understanding of food that they are not familiar with or they have not even seen before. That is where food recognition helps. It allows users from a picture of a food to identify the food name, food type, food ingredients, food components. By allowing people to identify the food they are eating, food recognition may facilitate dietary assessment, calorie intake estimation, nutrient analysis and eating habits analysis (Anthimopoulos et al. 2014).

Due to the rapid spreading of smartphones, food recognition using this tool is more popular. Thus, there is an increasing interest in applying food recognition to mobile environments to enable mobile food recognition. This has the advantage that detailed information on food can be obtained in real time (Bettadapura et al. 2015).

Decision making

Food information is also used for decision making. At the level of the general population, food information is used to choose not only palatable food but to make trade-offs between desirable food and the healthy ones. At the level of decision makers, many use cases are found. Food information is used to sensitize the population on the right eating behavior. Information on the geographic distribution of people suffering from malnutrition or people suffering from diseases caused by malnutrition may be used by deci-

sion makers to make awareness and take action such as the distribution of food supplements to children. Information on food traceability may help public health authorities to identify the cause of foodborne pathogens and outbreak. They can thus aid in the traceability of investigations, such as tracing the contaminated foods. A good example is the use of information on the relation between alcoholism and accidents, which allow policy makers to make alcohol driving laws.

Research

Food information engineering is a multidisciplinary field involving research fields such as food chemistry, food microbiology, sociology, medicine, nutrition, agronomy, environmental science and pollution, computer science, etc. Researchers from these domains will address different types of research problems. Researchers in medicine may be interested in the association between food and some diseases, the association between malnutrition and some diseases, etc. Researchers in food science and nutrition, food chemistry are generally interested in the composition of food in general and how to cope with this composition to provide a healthy diet to people. Toxicology research could be done to identify ingredient combinations or processes that can be the key factor of foodborne illness. Researchers in sociology may be interested in the food culture to understand the factors that influence the eating habits of the population, understand recipe popularity, etc. Researchers in environmental science and pollution may be interested in the detection and analysis of food poisoning caused by the pollution of regionally popular food. Researchers in agriculture may be interested in improving productivity and quality of crops.

These researches are supported by research in computer science in general. Research in software engineering can study how to build systems for following people's diet plan such as food logs or nutritional agenda (Metwally et al. 2021). This research produces a lot of data that is almost impossible to process manually. In addition, given the multifacet (agriculture, medicine, environment, etc.) and multi-modal (image, text, video) nature of food information, these systems do not only consider datasets individually, but also multiple datasets jointly. Thus, AI research will focus on the processing of these data to build AI systems. For instance, AI researchers in collaboration with domain researchers will provide by the means of symbolic knowledge representation a unified and standardized conceptual terminology and their relation to allow all these researchers to share knowledge.

FUTURE DIRECTION

Given the large spectrum that food information engineering covers, considerable effort is required to tackle the challenges and open issues. Existing review raised several future directions¹⁵ such as the construction of multi-modal food KG, the construction of large food image datasets, the development of large-scale robust food recognition system, food recommendation for nutritional health. In addition, during this work, we found that:

- In developing countries, food datasets such as food image datasets, recipe datasets, are still missing. The direct consequence is the absence of AI-based systems such as food recognition, food recommendation systems;
- Many food composition tables in developing countries are still in PDF format or in databases. However, the semantification of these data will improve their access to different stakeholders;
- Despite the multimodal nature of food information, the majority of food KG are in verbal mode, making them incomplete;
- The exploration of the relationships between diets and diseases should be considered. By constructing food KG with food composition and diseases instances, one can be able to analyze connections between diseases and food;
- Providing food information to different stakeholders according to their profile is a real challenge given that user profiles are diverse and may change. In the particular case of developing countries, several applications such as food log, food agenda, are still missing;
- Even if the multimodal nature of food information suggests the use of neurosymbolic techniques for processing and representing this information, this is currently rare;
- The large number of food information engineering research papers scattered on the Internet poses the problem of access to these scientific knowledge by the research community.

Providing responses to these issues will be a real added value in this domain.

DISCUSSION AND CONCLUSION

This paper presents how food information is collected, organized, processed, stored and used. To this end, we used ORKG templates to define conceptual models for describing several research problems, ORKG contributions for describing research papers based on the templates created, ORKG comparison tables for comparing research

contributions and ORKG smart reviews for reviewing the different aspects of food information engineering. All these resources are organized in the “Food Information Engineering” observatory. We found that food information is collected manually or automatically. Automatic collection and processing of food information relies on statistical, connectionists, symbolic or neurosymbolic techniques. Once collected, food information can be organized using tabular data representation schema, symbolic, connectionist or neurosymbolic AI techniques. Even if neurosymbolic AI have shown promising results in many domains, we found that this approach is rarely used in the domain of food information engineering. The diffusion of food information is provided according to the stakeholder profile. While the ones skilled in information management use statistical tools and specialized support, the others, mainly in the general population, use mass media and mobile phones. Information is provided to these people so as to be easily accessible and understandable.

Beside this work, we also documented related surveys using ORKG (Jiomekong 2023d). A large number of existing reviews consider mainly food recommendation systems. These systems aim at recommending food, restaurant, food ingredients, and so forth to people according to several criteria such as geographical region, food preference, health status. They propose the processing of food information using AI algorithms to predict food ingredients, palatable foods, healthy foods, restaurants. Existing review on the methodologies and tools for collecting food information is virtually nonexistent. Manual and automatic acquisition of food information is mentioned in several works without enough details on the methodologies, methods, and tools used. The processing and organization of food information using symbolic knowledge such as food classification systems, food ontologies, food knowledge graphs are proposed only in a few reviews. However, a lot of reviews propose the processing and organization of food information (mainly food images) using connectionists models. These models are used therefore for food image recognition, food recommendation, food calorie estimation, and so forth. Almost all review papers considered in this research highlighted the use of food information. It should be noted that unlike existing work, this work aims to cover the different aspects of food information engineering and to be linked to dynamic resources which will be continuously updated with new research papers and new contributions.

Even if the multimodal nature of food information suggests the use of neurosymbolic techniques for processing and representing these information, this is currently very rare. This paper in addition to presenting how food information is collected, organized and stored, also presents how neurosymbolic AI may be used for these purposes.

Thus, future direction of research in food information engineering may consist of carrying out the use of neurosymbolic AI for the collection, processing and organization of food information.


ACKNOWLEDGMENTS

This work was supported by the second and third ORKG curation grant provided by TIB-Leibniz Information Centre for Science and Technology. We are grateful to the ORKG team and all the grantees for their support, questions, remarks and suggestions on improvements.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict.

ORCID

Azanzi Jiomekong  <https://orcid.org/0000-0002-8005-2067>

Vogt Lars  <https://orcid.org/0000-0002-8280-0487>

ENDNOTES

- ¹ <https://orkg.org/>
- ² <https://orkg.org/stats>
- ³ Adding paper wizard: <https://orkg.org/add-paper>
- ⁴ <https://orkg.org/template/R677398>
- ⁵ <https://orkg.org/add-comparison>
- ⁶ <https://orkg.org/reviews>
- ⁷ <https://orkg.org/class/C25001>
- ⁸ <https://orkg.org/class/C34009>
- ⁹ <https://orkg.org/property/P62119>
- ¹⁰ <https://orkg.org/property/P62002>
- ¹¹ <https://orkg.org/template/R182248>
- ¹² https://orkg.org/observatory/Food_Information_Engineering
- ¹³ <https://www.yummly.com/>
- ¹⁴ <https://www.tripadvisor.com/>
- ¹⁵ <https://orkg.org/comparison/R691972>

REFERENCES

- Alazzai, W. K., B. Sh. Z. Abood, H. M. Al-Jawahry, and M. K. Obaid. 2024. "Precision Farming: The Power of AI and IoT Technologies." *E3S Web of Conferences* 491: 04006.
- Alert, A. E., J. H. Hollis, L. M. Lanningham-Foster, and M. Wolff. 2022. "The Influence of a Nutrition Recommendation on Food Choices in a Virtual Fast-Food Restaurant." *Current Developments in Nutrition* 6: 473–473.
- Anthimopoulos, M. M., L. Gianola, L. Scarnato, P. Diem, and S. G. Mougiakakou. 2014. "A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model." *IEEE Journal of Biomedical and Health Informatics* 18(4): 1261–71.
- Auer, S., A. Oelen, M. Haris, M. Stocker, J. D'Souza, K. Eddine Farfar, L. Vogt, M. Prinz, V. Wiens, and M. Y. Jaradeh. 2020. "Improving Access to Scientific Literature With Knowledge Graphs." *BIBLIOTHEK – Forschung und Praxis*, 44(3): 516–29.
- Bakr, M., S. Abdel-Gaber, M. M. Nasr, and M. Hazman. 2022. "DenseNet Based Model for Plant Diseases Diagnosis." *European Journal of Electrical Engineering and Computer Science*, 6(5): 1–9.
- Bettadapura, V., E. Thomaz, A. Parnami, G. D. Abowd, and I. Essa. 2015. "Leveraging Context to Support Automated Food Recognition in Restaurants." In *2015 IEEE Winter Conference on Applications of Computer Vision* 580–87.
- Bitá, A. I. G., A. Benjamin, T. Omgba, A. Agbornkwai, J. Eyong, and M. Sanou. 2020. "Qualitative Analysis of the Determinants of Malnutrition Among Children Under-five in Households in the Mayo-Danay Division, Cameroon." *Budapest International Research in Exact Sciences (BirEx) Journal* 3: 755–66.
- Brintha, N. C., P. Nagaraj, A. Tejasri, B. V. Durga, M. T. Teja, and M. N. Kumar. 2022. "A Food Recommendation System for Predictive Diabetic Patients using ANN and CNN." *2022 7th International Conference on Communication and Electronics Systems (ICCES)* 1364–71.
- Chen, Y., A. Subburathinam, C. H. Chen and M. J. Zaki. 2021. "Personalized Food Recommendation as Constrained Question Answering over a Large-scale Food Knowledge Graph." *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* 544–52.
- Connor, F. L., S. Salvatore, E. D'Auria, M. E. Baldassarre, M. Acunzo, G. D. Bella, I. Farella, S. Sestito, and L. Pensabene. 2022. "Cows' Milk Allergy-Associated Constipation: When to Look for It? A Narrative Review." *Nutrients* 14.
- D'Auria, E., E. Penderzza, and G. V. Zuccotti. 2020. "Personalized Nutrition in Food Allergy: Tips for Clinical Practice." *Frontiers in Pediatrics* 8.
- FAO. 2022. "FAO/INFOODS Food Composition Databases." <https://www.fao.org/infoods/infoods/tables-and-databases/faoinfoods-databases/en/>. [Online; accessed 2023-09-16].
- Goel, A. 2022. "Looking Back, Looking Ahead: Symbolic versus Connectionist AI." *AI Magazine* 42(4): 83–85.
- Greenfield, H., and D. A. Southgate. 2003. *Food Composition Data: Production, Management, and Use*. Food & Agriculture Org.
- Jiomekong, A. 2022a. "Food Information Processing." <https://orkg.org/review/R640411>. [Online; accessed 2023-09-16].
- Jiomekong, A. 2022b. "Food Knowledge Graphs." <https://orkg.org/review/R273427>.
- Jiomekong, A. 2022c. "Food Ontologies." <https://orkg.org/review/R273392>.
- Jiomekong, A. 2022d. "Using of Food Information." <https://orkg.org/review/R640415>. [Online; accessed 2023-09-16].
- Jiomekong, A. 2023a. "Collecting Food Information." <https://orkg.org/review/R609854>. [Online; accessed 2023-09-16].
- Jiomekong, A. 2023b. *Food Information Engineering: A Systematic Literature Review*. volume 37, 15441–15441.
- Jiomekong, A. 2023c. "Food Information Organization." <https://orkg.org/review/R640407>. [Online; accessed 2023-09-16].
- Jiomekong, A. 2023d. "Related Work on Food Information Engineering." <https://orkg.org/review/R642258>. [Online; accessed 2023-10-30].
- Jiomekong, A., S. Auer, and A. Oelen. 2024. "Linked Open Literature Review Using the Neuro-Symbolic Open Research Knowledge Graph." In *Companion Proceedings of the ACM on Web Conference 2024*, 1015–18. New York, NY, USA: Association for Computing Machinery.
- Jiomekong, A., G. Camara, and M. Tchuente. 2019. "Extracting Ontological Knowledge from Java Source Code Using Hidden Markov Models." *Open Computer Science* 9(1): 181–99.

- Jiomekong, A., and S. Tiwari. 2024. "An Approach Based on Open Research Knowledge Graph for Knowledge Acquisition from Scientific Papers." *The Electronic Library* 42(3): 413–42.
- Jiomekong, A., M. Uriel, H. Tapamo, and G. Camara. 2023. "Semantic Annotation of TSOTSATable Dataset." In *SemTab@ISWC*, 15–20.
- Liu, J., and X. Wang. 2021. "Plant Diseases and Pests Detection Based on Deep Learning: A Review." *Plant Methods* 17.
- Marinello, F., X. Zou, Z. Liu, X. Zhu, W. Zhang, Y. Qian, Y. Li, et al. 2023. "The Path to Smart Farming: Innovations and Opportunities in Precision Agriculture." *Agriculture*.
- Metwally, A. A., A. K. Leong, A. Desai, A. Nagarjuna, D. Perelman, and M. Snyder. 2021. "Learning Personal Food Preferences via Food Logs Embedding." In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2281–86. Los Alamitos, CA: IEEE Computer Society.
- Min, W., S. Jiang, L. Liu, Y. Rui, and R. Jain. 2019. "A Survey on Food Computing." *ACM Computing Surveys* 52(5).
- Min, W., C. Liu, L. Xu, and S. Jiang. 2022. "Applications of Knowledge Graphs for Food Science and Industry." *Patterns* 3(5): 100484.
- Pyingkodi, M., K. Thenmozhi, M. Karthikeyan, K. Chitra, N. Blessing, and S. Kumar. 2022. "Fruits Quality Detection using Deep Learning Models: A Meta-Analysis." *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 1–8.
- Nasar, Z., S. W. Jaffry, and M. K. Malik. 2018. "Information Extraction from Scientific Articles: A Survey." *Scientometrics* 117(3): 1931–90.
- Owino, V. O. 2019. *Challenges and Opportunities to Tackle the Rising Prevalence of Diet-Related Non-Communicable Diseases in Africa*. volume 78, 506–12. Cambridge University Press.
- Pan, S., L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu. 2023. "Unifying Large Language Models and Knowledge Graphs: A Roadmap." arXiv:2306.08302.
- Ponka, R., E. Fokou, E. Beaucher, M. Piot, and F. Gaucheron. 2016. "Nutrient Content of Some Cameroonian Traditional Dishes and Their Potential Contribution to Dietary Reference Intakes." *Food Science & Nutrition* 4: 696–705.
- Shah, P. K., C. Perez-Iratxeta, P. Bork, and M. A. Andrade. 2003. "Information Extraction from Full Text Scientific Articles: Where are the Keywords?" *Bmc Bioinformatics* 4: 20.
- Stocker, M., A. Oelen, M. Y. Jaradeh, M. Haris, O. A. Oghli, G. Heidari, H. Hussein, et al. 2023. "FAIR Scientific Information with the Open Research Knowledge Graph." *FAIR Connect*.
- WHO. 2023. "Malnutrition."
- Zhang, Y., W. Zhang, J. Yu, L. He, J. Chen, and Y. He. 2022. "Complete and Accurate Holly Fruits Counting Using YOLOX Object Detection." *Computers and Electronics in Agriculture* 198: 107062.
- Zulaika, U., A. Gutiérrez-Fandiño, and D. L. de Ipiña. 2018. "Enhancing Profile and Context Aware Relevant Food Search through Knowledge Graphs." In *International Conference on Ubiquitous Computing and Ambient Intelligence* 2(19) 1228. MDPI.

How to cite this article: Jiomekong, A., A. Oelen, S. Auer, L. Anna-Lena, and V. Lars. 2024. "Food information engineering." *AI Magazine* 45: 338–53. <https://doi.org/10.1002/aaai.12185>

AUTHOR BIOGRAPHIES

Azanzi Jiomekong is assistant professor in computer science in Cameroon and guest researcher at TIB. Since 2012, he has made important contributions to the application of Artificial Intelligence in health care in Africa. In 2023, his work on the development of an ontology for tuberculosis surveillance system in Cameroon received the best paper award at the Knowledge Graph and Semantic Web conference.

Allard Oelen is a postdoctoral researcher at TIB and the frontend lead for the Open Research Knowledge Graph. His interests include human–computer interaction, specifically human-AI collaboration for knowledge graph creation. This involves researching, designing, and developing user interfaces that enable users to produce and consume semantic knowledge.

Soren Auer is professor of data science and digital libraries at Leibniz Universität Hannover and the director of the TIB. He has made important contributions to semantic technologies, knowledge engineering, and information systems. He is co-author of over 200 peer-reviewed scientific publications and received several awards, including an ERC Consolidator Grant from the European Research Council, a SWSA 10-year award, the ESWC 7-year Best Paper Award, and the OpenCourseware Innovation Award.

Lorenz Anna-Lena is the community manager of the Open Research Knowledge Graph. She earned her PhD in theoretical physics in 2021 and is working for TIB - Leibniz Information Centre for Science and Technology since 2022. She is responsible for community building and training in the research data management project NFDI4DataScience.

Vogt Lars is the acting head of the area "Curation and Community Building" of the Open Research Knowledge Graph at the TIB Leibniz Information Centre for Science and Technology University Library. He has a PhD in evolutionary biology and is an ontology engineer with more than 15 years of experience in the field of ontologies, semantic modeling, and semantic interoperability.