

Enhancing Profile and Context Aware Relevant Food Search through Knowledge Graphs [†]

Unai Zulaika ^{*‡}, Asier Gutiérrez [‡] and Diego López-de-Ipiña [‡]

DeustoTech—University of Deusto, Avda. Universidades 24, 48007 Bilbao, Spain; asier.gutierrez@opendeusto.es (A.G.); dipina@deusto.es (D.L.-d.-I.)

* Correspondence: unai.zulaika@deusto.es

† Presented at the 12th International Conference on Ubiquitous Computing and Ambient Intelligence (UCAmI 2018), Punta Cana, Dominican Republic, 4–7 December 2018.

‡ These authors contributed equally to this work.

Published: 25 October 2018



Abstract: Foodbar is a Cloud-based gastroevaluation solution, leveraging IBM Watson cognitive services. It brings together machine and human intelligence to enable cognitive gastroevaluation of “tapas” or “pintxos”, i.e., small miniature bites or dishes. Foodbar matchmakes users’ profiles, preferences and context against an elaborated knowledge graph based model of user and machine generated information about food items. This paper reasons about the suitability of this novel way of modelling heterogeneous, with diverse degree of veracity, information to offer more stakeholder satisfying knowledge exploitation solutions, i.e., those offering more relevant and elaborated, directly usable, information to those that want to take decisions regarding food in miniature. An evaluation of the information modelling power of such approach is performed highlighting why such model can offer better more relevant and enriched answers to natural language questions posed by users.

Keywords: knowledge graphs; relevant search; recommendation systems; data models; inference; software architectures

1. Introduction

Currently, many of us follow social opinion when choosing where to go for food. Particularly, when we go to a unknown environment. Google Places (<https://developers.google.com/places/web-service/intro>) and TripAdvisor (<https://www.tripadvisor.com/>) are two widely used services for this purpose. Still, those apps are more catered for recommending restaurants based on user opinions rather than recommending specific dishes or food specialties at different places in a geographical range, still taking into account social opinion and ratings. There is, indeed, the need to advance on food recommendation systems rather than restaurant, cafe or bar recommendation systems which do not only cater with vague explicit requests from users, e.g., “I feel like having something fishy” but also take into account the user’s profile, e.g., their dietary restrictions like “I am allergic to corn” and their context, e.g., “It is Monday morning and I am in the city centre”.

Big corporations such as Amazon, Apple, Google, Microsoft or IBM are progressing towards offering fully fledged more elaborated answers to users’ requests/queries. Assistants like Cortana, Siri, Now or Alexa are good examples of services where user context, profile together with their explicit interaction, often through natural interfaces, like voice, get into play. Often, behind those assistants or agents, there is a sophisticated data model which enables to go beyond being able to return ranked lists of results related to the terms used in. Indeed, the latest trend is that those assistants return elaborated natural language answers already correlating or including links to relevant

related information. Google Knowledge Graph [1] is a well-known solution which uses this approach, thus Google is progressively turning from a search engine into a Question and Answer Engine. However, despite the interest on the Knowledge Graph as a new way of modelling and exploiting knowledge, the bibliography on this topic is limited and only few mentions in industry blogs are referring to examples where Knowledge Graphs are being used in other domains that searching, e.g., in Industry 4.0 by SIEMENS (<https://paul4innovating.com/2017/12/29/as-we-enter-2018-we-will-need-knowledge-graphs/>).

Foodbar is a Cloud-based gastroevaluation solution, leveraging IBM Watson cognitive services, bringing together machine and human intelligence to enable cognitive gastroevaluation of “tapas” or “pintxos”, i.e., small miniature bites or dishes. The foodbar solution matchmakes users’ profiles, preferences and context against an elaborated model of user and machine generated information about food facts. The information associated to a pintxo combines heterogeneous information, with diverse degrees of veracity, provided either by a bar owner or by their customers, e.g., photos, comments and ratings, with information extracted from external data repositories, e.g., recipes or calories contents of different food items. The system, on one hand, enhances recommendation about miniature food with cognitive features, e.g., a chat bot, sentiment analysis of comments or automatic comment summarization. On the other hand, it offers to restaurant managers descriptive analytics services so that they can understand why and when some pintxos are working well and others don’t. That way, decision making can be facilitated across distinct pintxo stakeholders.

This paper concentrates on describing the internal knowledge modelling approach chosen for foodbar. It advocates that domain-focused knowledge graphs are the way to offer more stakeholder satisfying knowledge exploitation solutions which can offer more relevant and elaborated, directly usable, information to those that want to take consumption or design/selling decisions regarding food in miniature. A rationale about the information modelling power of such an approach is performed highlighting why such model can offer better more relevant and enriched answers to natural language questions posed by users. The structure of the paper is as follows. Section 2 describes some related work. Section 3 explains the foodbar solution, its current approach and how can be enhanced by adopting knowledge graphs. Section 4 explains the foodbar knowledge graph architecture and different core processes to make usable such model in the long run, namely, population, inference or knowledge generation and quality assurance, and exploitation through queries. Section 5 shows the validation of foodbar by showing its current configuration and ongoing work on transforming it into a more sophisticated cognitive system. Finally, Section 6 draws some conclusions and outlines future work.

2. Related Work

Knowledge construction aims to create a resource where information from different fields is gathered. Different approaches in the form of a Knowledge Graph or Knowledge Base have been done for that. Those examples may include Google’s Knowledge Graph (<https://www.google.com/intl/bn/insidesearch/features/search/knowledge.html>), Knowledge Vault [2], NELL [3] or DeepDive [4]. In fact, the Linked Open Data itself can be considered as a huge knowledge base. Nevertheless, a big drawback for such wide perspectives is the lack of precision for a specific domain. They are meant to cover and respond general search queries which is not the case for foodbar where deep but only food domain questions are done.

Domain related solutions have also been proposed, especially with commercial proposes. Those include Amazon’s product graph (http://kbcom.org/speaker_slides/ProductGraph_KBCOM.pdf) or Siemens’ Industrial Knowledge Graph (https://indico.cern.ch/event/669648/contributions/2838194/attachments/1581790/2499984/CERN_Open_Lab_Technical_Workshop_-_SIEMENS_AG_-_FISHKIN_-_11-01-2018.pdf). For example, in [5], a Knowledge Graph is used as solution for combating human trafficking. These approaches share a need for specific kinds of data and deep domain understanding, like in foodbar project. Unlike heterogeneous knowledge graphs where

information extraction techniques and data sources are similar, this homogeneous knowledge graphs require specific techniques and data sources to be valuable. Furthermore, in order to provide relevant results, data must be rich and detailed.

Relevance is the practice of improving search results for users by satisfying their information needs in the context of a particular user experience [6].

Regarding relevant search techniques, new research [7] suggests that successful text-based personalization algorithms perform significantly better than explicit relevance feedback where the user has fully specified the relevant documents. Knowledge graphs provide rich contexts where personalization algorithms can perform well.

3. Foodbar Solution

In this section we describe the current approach for foodbar solution and we present the reasons that drive us to adopt a Knowledge Graph-based solution using different information retrieval techniques.

3.1. Current Approach

Foodbar is a cognitive and social intelligence application for enhancing competitiveness of miniature food sector. The core of the application resides in users' ratings and opinions about pintxos. Opinions are sentimentally analyzed to get the feeling they provoke. More over, those ratings and opinions are used for a recommendation algorithm based on users' preferences. A conversational agent is also one of the main features. This agent recommends a pintxo after gathering some information through a conversation with the user. Besides, an analytics dashboard is provided to pintxo bar owners. The dashboard presents statistics about pintxos: most liked ones, feelings they generate and so on. Otherwise, foodbar behaves as an ordinary application. It follows the standard client-server architecture, has some CRUD operations, user settings and basic search functions.

All those functionalities are dependant of a document based NoSQL database, i.e., MongoDB. The data model for this database is the shown in Figure 1.

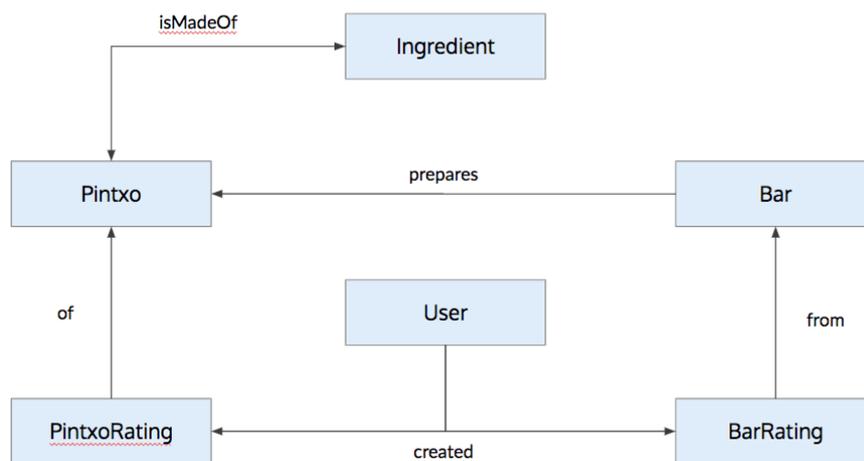


Figure 1. Diagram for foodbar data model.

The model has three main entities: *Pintxo*, *Ingredient* and *Bar*. *Ingredients* are the entities that make up pintxos which are generic types of recipes, i.e., bread, egg and potato are ingredients for a potato omelette (Spanish omelette) pintxo. However, each *Bar* (business) has a list of pintxos with custom characteristics for that specific bar. This is because even though two bars can have the same recipe or

pintxo, those are different instances. For instance, the potato omelette from bar A has some distinctive features to the potato omelette of bar B. Additionally, there are other entities for *users*, *ratings* and *conversations* with the conversational agent.

Three main improvements have been identified for the application in order to make it “more cognitive”, i.e., closer to the way humans ask and response questions:

- *Deep queries*: it should be able to understand and provide good results for complex queries that require a high level of context understanding.
- *Flexible data modeling*: a system to facilitate data and relationship modeling, furnishing a soft structure to work with recipes, ingredients, diets, food families etc.
- *Entity validation*: one of the main challenges is the veracity level of data. In order to avoid overloading users with cumbersome validation duties, automatic validation techniques must be applied.

Foodbar’s cognitive skills are optimized by creating a knowledge graph for the purpose.

3.2. Benefits of a Knowledge Graph in Foodbar

Modeling and representing food is a very hard task even for a human person. Relationships between ingredients, recipes or diets are difficult to identify and contextualize to make them a valuable asset. The main objective of foodbar is to provide a cognitive and intelligent experience to the user. This could be achieved by providing relevant responses built upon contextualized information. For that, it is necessary to create a structure of knowledge which should contain every possible data that might add value to food related queries. The best resource that can be created for the task is a knowledge graph where complex relationships between entities can be properly represented, and context can be taken into account.

Knowledge graphs are knowledge bases to enhance search results with semantic information collected from a wide range of resources. Thanks to their graph structure, as stated in [8], they have a sheer performance increase when dealing with connected data versus relational databases and NoSQL stores. Furthermore, graphs are naturally additive, feature that perfectly fits with the dynamic and growing domain of food and its context. Most importantly, they store data in a connected way. This feature is key to find semantic and context dependencies between entities, a process that would otherwise involve low efficiency processes in SQL and other NoSQL databases. For example, consider the food domain for ‘cheese’ shown in Figure 2.

As this variably-structured and densely connected data network grows, there will be huge time increases in searches for non graphs stores. In contrast, graph stores can handle data and relationships growth properly while increasing their knowledge and providing preciser answers. More sophisticated queries would be possible from the user perspective, such as:

- Give me pintxos that will make me feel good.
- Show me a non dairy pintxo that I will love.
- Why are the top three pintxos that popular?
- Recommend me pintxos that do not make me fat.
- Which pintxo should I take today?

This type of queries can only be addressed in a relevant and efficient way by building a context aware, cognitive knowledge system, which can be modeled in a graph structure as proposed by this work. Moreover, some network related algorithms such as PageRank [9] or Louvain [10] are of major interest for foodbar project and can be easily applied into a graph structure.

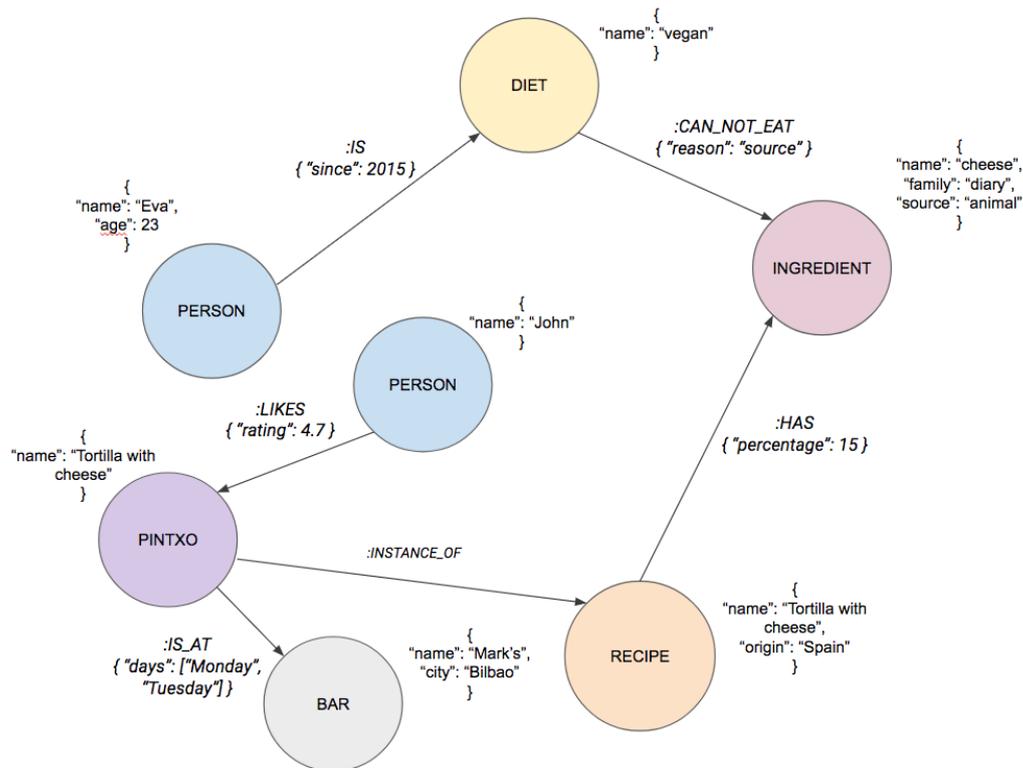


Figure 2. Graph data model and example for 'cheese ingredient'.

4. Foodbar Knowledge Graph

In this section, the approach for foodbar knowledge graph is presented. We place knowledge graphs within cognitive computing [11] research domain. They are resources where relevant data for improving human decision-making and finding behavioural patterns is stored. This is a key feature for foodbar as it does aim to be an intelligent application.

4.1. FKG Architecture

The proposed architecture for the FKG is shown in Figure 3:

It contains several components that work together to create a search and query infrastructure that behaves in a cognitive way and provides relevant responses to ensure the best possible user experience. Each of those components have a specific task:

- A** *Foodbar knowledge graph*: this is where knowledge for foodbar is stored. It holds different facts that describe the food domain, e.g., pintxo details, ingredients, opinions or ratings, and relates them to foodbar's users, bars, points of interest, cultural facts and other information that enriches the graph.
- B** *Data discovery*: the main task of this component is to continuously (or regularly at least) extract food related information from different data sources. There is further information in the following Section 4.2.
- C** *Foodbar DB*: it is the underlying database for foodbar application.
- D** *Contextualizer*: it is a service for cleansing, augmenting (adding new relationships) and merging data from FKG. It is a service which pursues the continuous optimization of the FKG.
- E** *Foodbar knowledge graph index*: it is a quick access document indexing system derived from the FKG. It provides a fast way to get data.

- F** *FKG API*: this component offers a service to query and search for foodbar domain.
- G** *Foodbar App*: application that makes use of FKG API to offer a cognitive experience to users.

Each of these components fulfill a role in the three main processes that have been identified for the project. Those processes, are explained in the next three sections.

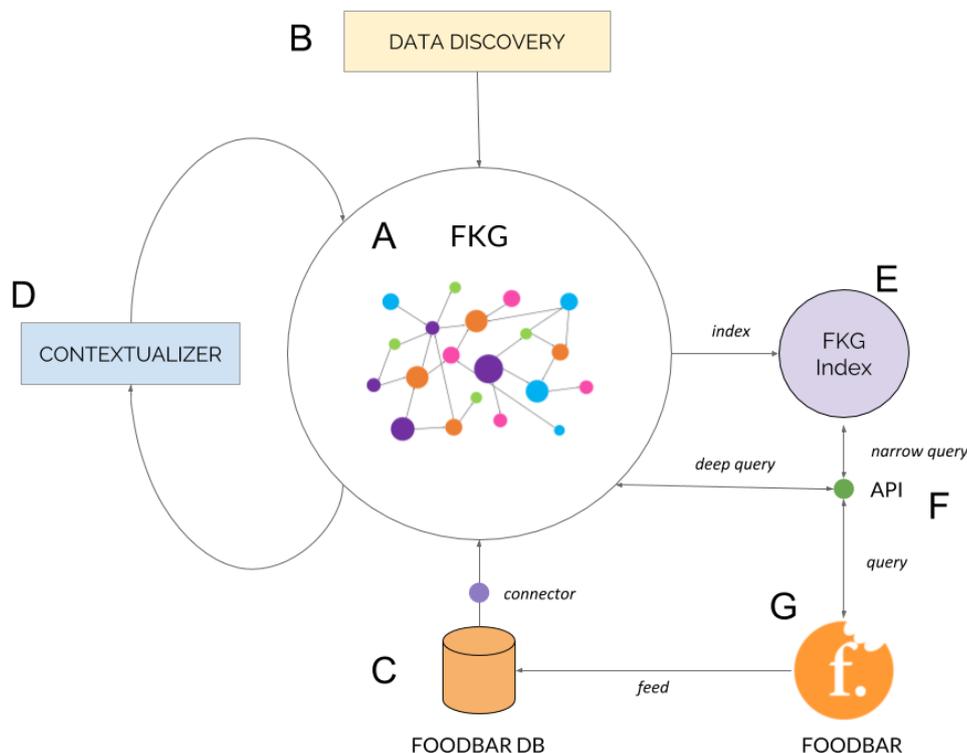


Figure 3. Architecture for foodbar knowledge graph.

4.2. Data Extraction and Population Strategy

Several techniques, for instance [2] or [12], have been suggested for automatic knowledge construction. Different approaches like YAGO [13], WikiData [14], DBpedia [15], ConceptNet [16], Reverb [17], DeepDive [4] or ProBase [4] have been proposed to construct a cross domain knowledge base. In contrast, FKG is tightly related to the food domain which enables to focus in specific goals when extracting data and populating the graph.

Two different types of data sources have been identified for the FKG, from the foodbar application itself and from outside foodbar. Referring to them by their component name in Figure 3, the first type of data source comes from the foodbar DB component, and the second from data discovery. Foodbar database holds information created by the application users, this is data like bars, geographic coordinates, pintxos, ratings, comments or habits. Most of this information adds context value to FKG and improves complex queries that involve user preferences. Moreover, the data discovery component is a service which crawls food domain related datasets from the web. This data is fundamental for the FKG to be able to have a wide amount of information and become a valuable resource to data coming from foodbar application. Considering that FKG does need a niche of concrete data, that belonging to food domain, our population strategy focuses on crawling data from two specific sources: WikiData and BEDCA (<http://www.bedca.net/>). WikiData has been selected given its amount of information and because it contains Cookbook (https://en.wikibooks.org/wiki/Cookbook:Table_of_Contents). Cookbook is a growing and world-wide collection of recipes translated into, or written in, English with links to other languages that has been identified to be an invaluable source

of information. BEDCA has been chosen due to the localized value of the information it contains for foodbar (being a Spanish database).

4.2.1. SPARQL Extraction

For populating WikiData a microservice is used. This is in charge of querying WikiData’s SPARQL [18] endpoint and inserting the results to FKG. The philosophy is to query about food domain entities and not other domain related data in order to avoid possible data noise [19]. For this task queries like the next one are performed.

```
#Get all subclasses of dairy products
SELECT ?item ?itemLabel ?Freebase_ID WHERE {
  # statement (semantic triple): ?item subclassOf dairy products
  ?item wdt:P279 wd:Q185217.
  # query wikibase service and retrieve information in English
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
  # if exists: retrieve Freebase ID for every item
  OPTIONAL { ?item wdt:P646 ?Freebase_ID. }
}
```

In the sentence above, wikibase (<https://www.mediawiki.org/wiki/Wikibase/en>) service is asked to retrieve all items, their labels and Freebase IDs for every entity *subclassOf* dairy products. Terms included in the the query are:

- *item*: entity in Wikidata.
- *itemLabel*: label for an entity, i.e., ‘Cheese’.
- *Freebase ID*: ID for the class in the Google Knowledge Graph, this is interesting in order to apply distant supervision techniques as the one presented in DeepDive [4]. In distant supervision, an already existing database is used, such as Freebase which was acquired by Google in 2014 to create the Google Knowledge Graph, or a domain-specific database, to collect examples for relations that want to be extracted.
- *wdt:P279*: represents *subclassOf* relation.
- *wd:Q18*: represents dairy product entity in WikiData.
- *wdt:P464*: represents *Freebase ID* relation.

A example result for the query can be observed in Table 1.

Once results have been retrieved, the microservice maps and connects responses into FKG. It is important to react to changes and updates of already imported contents. MediaWiki, provides an API where every change in Wikimedia Foundation is recorded (<https://www.mediawiki.org/w/api.php?action=feedrecentchanges>). By frequently checking this, it is possible to find changes in the data queried to feed the FKG. Therefore, there is a service which daily checks for those changes, and in case of finding updates in food domain elements it does introduce them into the graph. This way, it is guaranteed that the FKG is up to date with a constantly evolving knowledge base like WikiData.

Table 1. Results for SPARQL query.

Item	ItemLabel	Freebase ID
wd:Q10943	cheese	m01nkt
wd:Q13233	ice cream	m01tv9
wd:Q13317	yogurt	m0cxn2

4.2.2. BEDCA Crawling

BEDCA is a network of public research centers, public administrations and institutions for developing and maintaining a Spanish Database of Food Composition. This database contains detailed information about food: numbers of vitamins, proteins, fat levels, etc. This information is a huge input for the graph because it provides data for crafting complex diets and health related recommendation among other relevant functions. It is important to also highlight the difficulty involved in obtaining this data from users. Besides, BEDCA does classify food in groups based on their source, for example cereals or vegetables. All data is translated and localized to Spanish cuisine which is very important for foodbar context as addressed in the introduction Section 1.

To obtain information from BEDCA website, a script in Python has been used. This script is really specific for the website and will not be reused for other purposes. Once crawled, all data goes into FKG. The website is not frequently updated which allows us to run the script every six months and look for changes in the database.

4.3. From Information to Knowledge

The Oxford dictionary defines knowledge as (<https://en.oxforddictionaries.com/definition/knowledge>):

Facts, information, and skills acquired through experience or education; the theoretical or practical understanding of a subject.

It is fairly easier to understand facts in a relatively small domain such as the food one. In this sense, FKG aims to be a clean and precise resource or encyclopedia for machines. A tool for machines to learn and become smarter in the food domain.

To achieve this purpose, there are some processes regarding data enhancement that must be done, namely (a) *Cleansing*; (b) *Inferencing* and (c) *Localizing*.

Cleansing is the process for finding possible duplicates, corruptions and removing unreliable data in the graph. In foodbar context, where users provided data is merged and linked with web extracted data that can be deceptive, establishing proper bridges/relationships and techniques to have correct information is a priority. For instance, if a user from foodbar feeds the FKG with a new recipe that has been recently created, there will be no prior registers for that recipe. Hence, FKG must not classify that recipe as untruthful. Besides, facts that are similar must be simplified. Another way to validate data is by having human people playing the role of “validators”. To achieve the validation it is necessary to engage users and gamification engines and rewarding systems are a great tool for that. This can help to keep reliable data in the graph.

Data *inferring* is about obtaining a third fact from two of them. The Contextualizer from Figure 3 is in charge of this task. Previous approaches have been presented for inferring, most of them based on OWL/RDF, for example HerMiT [20]. Although this kind of reasoners have been widely used in RDF structures, they can not be directly applied to labeled property graphs, as the one used by popular graph databases such as Neo4j. Foodbar uses Cypher, a graph query language for Neo4j which is the database that will be holding the FKG. Cypher allows making semantically enhanced queries which carry out inference in real time. For that, it is needed to create a custom endpoint which maps inference rules for given data. An example for this can be found is in the following repository (<https://github.com/jbarrasa/neosemantics>).

Localizing is about translating and contextualizing data in the environment of a culture or region. For example, in the USA the term ‘pancake’ is the most common one, but in some states they are called ‘flapjack’ or ‘silver dollar’. Due to the singularities that the foodbar project faces, this is a key feature. In fact, the foodbar project was born in the Basque Country (a region with its own language from northern Spain), where lots of food domain related terms come from Basque language itself.

4.4. Querying FKG

Given their complexity, queries issued to FKG are classified into two types: deep and narrow. The API service from foodbar is in charge of classifying query types and responding them using the underlying proper answer generation resource.

On the one hand, *narrow queries* are those that can be easily done, i.e., a single document query. They can be processed and responded really fast as a result of being easy to look up items in the FKG Index (see Figure 3). Currently, the FKG Index used is an Elasticsearch instance which holds FKG derived documents. A narrow query for that is, for example, which the most liked pintxo with cheese is.

```

1 {
2   "query": {
3     "term": {
4       "pintxos.pintxo.ingredients.name": "queso"
5     }
6   },
7   "sort": [
8     {
9       "pintxos.rating": {
10        "order": "desc"
11      }
12    }
13  ]
14 }

```

On the other hand, *deep queries* are complex queries that can not be responded by using the FKG Index. These ones are directly queried to the FKG and require a high level of cognition and relevancy in search. Those, responses include many relationships, what makes asking a graph structure more efficient. For instance, which will be the most eaten pintxo tomorrow.

5. Validation

The current running instance of foodbar data model (see Figure 1) contains exclusively user-generated data. For that, a group of 23 users where requested to visit the same 3 bars in Bilbao, taste 6 pintxos and offer elaborated comments and ratings about them that were registered through a Google Form as shown in Figure 4.

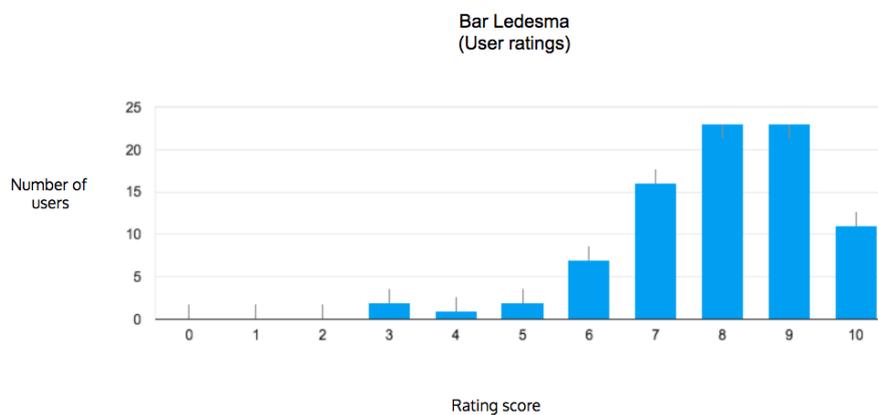


Figure 4. Ratings gathered from users.

As result of processing the resulting questionnaires’ dataset, we generated a small but high quality dataset that has allowed us to test the recommendation capabilities of foodbar, either selecting the option to receive a recommendation by clicking a button from the interface or asking the chatbot to give a recommendation. The small group of 23 users has tested the recommendation capabilities and have a favourable opinion regarding the acceptance as relected by Technology Acceptance (TAM (https://en.wikipedia.org/wiki/Technology_acceptance_model)) questionnaire.

Currently, foodbar development is focused on the implementation of microservices to achieve data discovery and contextualize the data gathered, thus progressing from data into connected knowledge. Our main challenge now is how to convert semantic triples obtained from WikiData into a Labelled Property Graph (LPG) model. A triple is a statement about semantic data in the form of subject–predicate–object expressions, while LPG can embed those expressions in entities and relationships in form of properties. Differences are evident when comparing Figures 2 and 5.

A good approach for adapting triples to LPG is crucial in order to ensure that Cypher inference explained in Section 4.3 works properly.

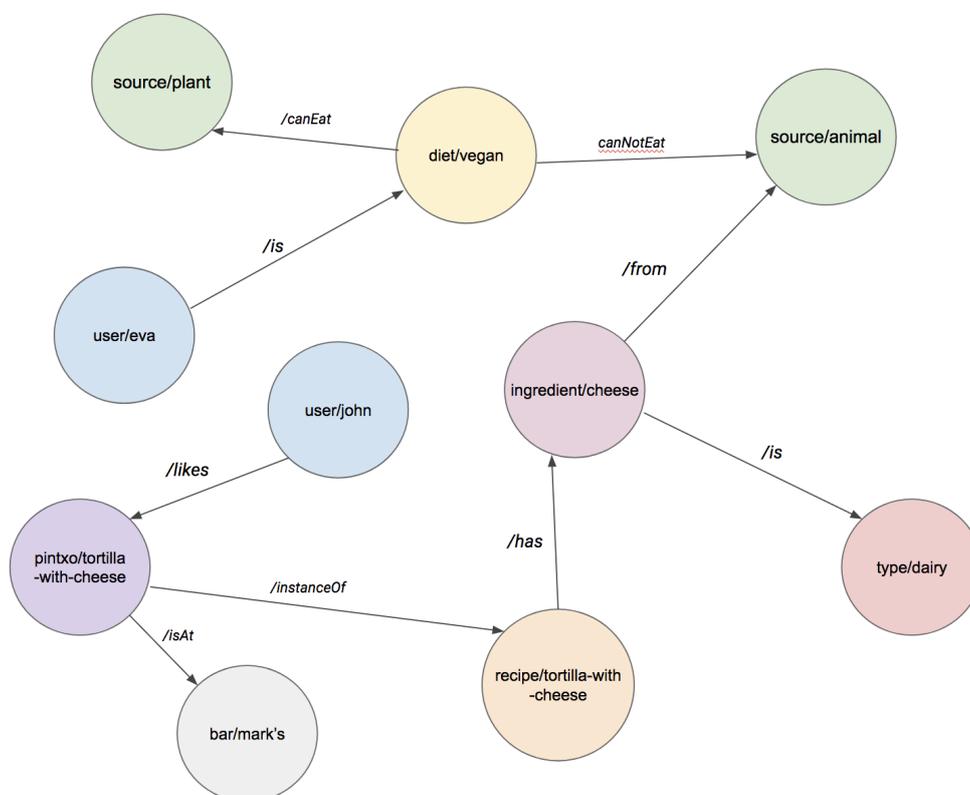


Figure 5. Example data from Figure 2 (some properties do not appear) represented in semantic triples.

6. Conclusions and Further Work

This paper presents a novel approach based on knowledge graphs to offer cognitive features when searching or being recommended about miniature food, i.e., pintxo, in a location and according to a given user profile or preferences. A graph structure is more suitable for a complex domain like the food one where relationships between data are paramount. Furthermore, due to the fact that relationships are first-class citizen, these structures are environments where recommendation systems perform well. Foodbar clearly benefits from this kind of structures. By adding new sources of data, and contextualizing it, it becomes a more intelligent solution.

The foodbar knowledge graph is a constantly optimizing entity. First, new data is being introduced into the graph both from foodbar application itself and from other data sources, currently WikiData and

BEDCA. Additionally, a process to contextualize and infer new information is continuously running on the graph to increase and improve results. FKG is therefore continuously aiming to augment its knowledge and become an even more precise and valuable resource to understand users and give them better and more elaborated relevant results.

Domain focused graphs require more precision in every aspect of the knowledge construction. Detailed and reliable data sources are needed and contextualizing methods must be accurate. Veracity plays an important role because the given results must satisfy users who expect high quality and in-depth answers.

The following further work activities will be considered:

- New data sources will be considered for enriching the FKG. Several knowledge bases that could be interesting have been identified, for example KBpedia (<http://kbpedia.com/>) which does itself contain information from various sources. Such potential integration demands personalized solutions to properly work into the graph and add valuable information.
- A hybrid approach for continuously enhancing the quality of the information modeled by the FKG will be explored, where machine learning and human intelligence, i.e., human intervention, through moderation or gamification interfaces might be explored.
- Seek new exploitation models for both machines and people to improve knowledge accessibility. Methods and models to create and maintain knowledge are key to enhance people's life, foodbar is a big opportunity to follow this track.
- Investigate methods for the purpose of more precisely localize information in a cultural way. We want the system to deeply understand subtle differences that may exist between different cultures.
- Keep up improving results in our recommendation system is one of the main tasks. It is vital to provide a cognitive experience to users and recommendations are the main contribution for that, being a field subject to continuous revision and improvement.
- Analytics are of huge relevance for business owners in foodbar. Retrieving valuable information from the FKG by applying graph algorithms is an important line of research.
- In order to allow other systems to integrate foodbar, we need to expose metadata about the graph. By doing this, third parties will be able to discover and properly understand the structure of the FKG and work with it. For this, adding a tool like HyperCat (<https://hypercatiot.github.io/>) is an interesting option.

Acknowledgments: This work has been possible thanks to IBM Shared University Research Award. We thank Luis Rico for conceiving the foodbar concept and Víctor López Fandiño, Izaskun Santiesteban and Elisa Martín Garijo from IBM Spain, for their continuous follow-up and advice relative to IBM Watson technologies.

References

1. Huynh, D.F.; Li, G.; Ding, C.; Huang, Y.; Chai, Y.; Hu, L.; Chen, J. Generating Insightful Connections between Graph Entities. U.S. Patent 20140280044A1, March 2013.
2. Dong, X.; Gabrilovich, E.; Heitz, G.; Horn, W.; Lao, N.; Murphy, K.; Strohmann, T.; Sun, S.; Zhang, W. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 601–610.
3. Mitchell, T.; Cohen, W.; Hruschka, E.; Talukdar, P.; Yang, B.; Betteridge, J.; Carlson, A.; Dalvi, B.; Gardner, M.; Kisiel, B.; et al. Never-ending learning. *Commun. ACM* **2018**, *61*, 103–115.
4. Niu, F.; Zhang, C.; Ré, C.; Shavlik, J.W. DeepDive: Web-scale Knowledge-base Construction using Statistical Learning and Inference. *VLDS* **2012**, *12*, 25–28.
5. Szekely, P.; Knoblock, C.A.; Slepicka, J.; Philpot, A.; Singh, A.; Yin, C.; Kapoor, D.; Natarajan, P.; Marcu, D.; Knight, K.; et al. Building and using a knowledge graph to combat human trafficking. In Proceedings of the International Semantic Web Conference, Bethlehem, PA, USA, 11–15 October 2015; pp. 205–221.
6. Turnbull, D.; Berryman, J. *Relevant Search: With Applications for Solr and Elasticsearch*; Manning Publications: Shelter Island, NY, USA, 2016.

7. Teevan, J.; Dumais, S.T.; Horvitz, E. Personalizing search via automated analysis of interests and activities. In Proceedings of the ACM SIGIR Forum, Ann Arbor, MI, USA, 12 July 2018; Volume 51, pp. 10–17.
8. Robinson, I.; Webber, J.; Eifrem, E. *Graph Databases: New Opportunities for Connected Data*; O'Reilly Media: Newton, MA, USA, 2015.
9. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Technical Report; Stanford InfoLab: Stanford, CA, USA, 1999.
10. De Meo, P.; Ferrara, E.; Fiumara, G.; Provetti, A. Generalized louvain method for community detection in large networks. In Proceedings of the 2011 11th International Conference on Intelligent Systems Design and Applications (ISDA), Córdoba, Spain, 22–24 November 2011; pp. 88–93.
11. Kelly, J., III; Hamm, S. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*; Columbia University Press: New York, NY, USA, 2013.
12. Grainger, T.; AlJadda, K.; Korayem, M.; Smith, A. The Semantic Knowledge Graph: A compact, auto-generated model for real-time traversal and ranking of any relationship within a domain. *arXiv* **2016**, arXiv:1609.00464.
13. Suchanek, F.M.; Kasneci, G.; Weikum, G. Yago: A core of semantic knowledge. In Proceedings of the 16th International Conference On World Wide Web, Banff, AB, Canada, 8–12 May 2007; pp. 697–706.
14. Vrandečić, D.; Krötzsch, M. Wikidata: A free collaborative knowledgebase. *Commun. ACM* **2014**, *57*, 78–85.
15. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. Dbpedia: A nucleus for a web of open data. In *The Semantic Web*; Springer: Berlin, Germany, 2007; pp. 722–735.
16. Liu, H.; Singh, P. ConceptNet—A practical commonsense reasoning tool-kit. *BT Technol. J.* **2004**, *22*, 211–226.
17. Fader, A.; Soderland, S.; Etzioni, O. Identifying relations for open information extraction. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Edinburgh, UK, 30 March 2011; pp. 1535–1545.
18. Prud, E.; Seaborne, A. SPARQL Query Language for RDF 2006. Available online: <http://www.w3.org/TR/rdf-sparql-query/> (accessed on 15 October 2018).
19. Pujara, J.; Augustine, E.; Getoor, L. Sparsity and Noise: Where Knowledge Graph Embeddings Fall Short. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 1751–1756.
20. Shearer, R.; Motik, B.; Horrocks, I. Hermit: A Highly-Efficient OWL Reasoner. *OWLED* **2008**, *432*, 91.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).