

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

CAMPUS DI CESENA
SCUOLA DI INGEGNERIA

Corso di Laurea Magistrale in Ingegneria e Scienze Informatiche

FOODON

Elaborato del corso di: Web Semantico

Report di:
LORENZO CHIANA

ANNO ACCADEMICO 2019–2020

Indice

1	Panoramica	1
1.1	Motivazioni	1
1.2	Ontologie e cibo	2
1.3	Il progetto FoodOn	2
1.4	Obiettivi	3
2	FoodOn	5
2.1	Ontologia	5
2.1.1	FoodOn e LanguaL	6
2.1.2	Gerarchie	8
2.1.3	Riutilizzo dell'ontologia	9
2.2	Struttura	12
2.3	Relazioni	18
3	Conclusioni	21

Capitolo 1

Panoramica

1.1 Motivazioni

Al giorno d'oggi l'innovazione digitale sta influenzando molti aspetti sanitari ed economici della produzione, distribuzione e consumo degli alimenti. Ad esempio l'Internet of Things (IoT) propone sempre più una visione di ambienti come quelli agricoli e industriali pervasi di sensori connessi costantemente alla rete, generando dati che possono essere utilizzati per aumentare la qualità del prodotto e del processo stesso. Nello specifico, in ambito agricolo e gastronomico, i dati generati da queste tecnologie possono essere utilizzati per aumentare la qualità degli alimenti e garantire la tracciabilità, riducendo al contempo costi, consumi e sprechi legati alle risorse.

Altro esempio le questioni legate alla sicurezza degli alimenti, all'autenticità e ai conflitti derivanti dalla protezione dei marchi e tutta la logistica sull'approvvigionamento e sulla distribuzione degli alimenti locali contro multinazionali possono e vengono analizzate anche con l'aiuto di set di dati e modelli specifici per gli alimenti.

Questo scambio di dati sempre più crescente ha messo in luce un problema fondamentale: pochi sottodomini terminologici sono stati standardizzati un po' in tutti i settori. La pleora di dizionari alimentari mantiene invisibili le informazioni sugli alimenti a causa della mancanza di interoperabilità, influenzando di conseguenza la tracciabilità degli alimenti, dei patogeni di origine alimentare, dei contaminanti e, più in generale, della qualità degli alimenti stessi. In passato si è già tentato di creare vocabolari alimentari applicabili a livello internazionale, ma ostacoli sia tecnici che linguistici

hanno scoraggiato un deposito globale per la catalogazione degli alimenti regionali e della loro composizione.

La combinazione dell'attuale infrastruttura Internet e dei progressi del web semantico rende ora interessanti soluzioni di tipo ontologiche.

1.2 Ontologie e cibo

Un'ontologia fornisce una teoria formale per un dominio di indagine che specifica il significato dei termini all'interno di un vocabolario e consiste in una struttura tassonomica gerarchica e suddivisa in assiomi che permettono di capire come le entità all'interno di un dominio sono correlate. Termini appropriati possono essere identificati e distinti da etichette ontologiche e sinonimi che includono nomi in multilingua o nomi specifici della regione, nonché identificatori e definizioni univoci accessibili a livello globale, evitando così l'uso di vocabolari ambigui. Le ontologie sono inoltre in grado di accogliere più gerarchie, spesso sotto forma di tassonomie, che possono agire con sfaccettature a maggior o minor livello di dettaglio mentre si naviga tra le gerarchie stesse. Un prodotto alimentare, ad esempio, può essere collegato a varie categorie di prodotti alimentari nazionali o internazionali in una sfaccettatura "tipo prodotto", così come gli ingredienti per mezzo di una sfaccettatura gerarchica di animali e/o piante "fonte di cibo".

Le ontologie ben progettate permettono di riutilizzare termini di altre ontologie già consolidate al fine di eliminare duplicati. Ciò permette l'integrazione di ontologie altrimenti disparate tra domini. Inoltre un'ontologia consente di migliorare un termine del vocabolario attraverso assiomi logici che un calcolatore può leggere e su cui può ragionare.

1.3 Il progetto FoodOn

FoodOn è un progetto guidato dal consorzio Hsiao Lab per costruire un'ontologia globale *farm-to-fork*, ossia un vocabolario per la descrizione alimentare standardizzato per supportare la ricerca, le applicazioni di consumo e quelle industriali. Al consorzio si sono poi rapidamente unite figure provenienti dall'Università della British Columbia, dal British Columbia Center for Disease Control Public Health Laboratory e dalla comunità OBO Foundry avendo esigenze parallele per l'agricoltura, l'analisi nutrizionale e la

ricerca sulle scienze alimentari. La proposta di FoodOn si basa sulla condivisione e promuove uno standard che nasce dalla congiunzione di domini ontologici coordinati e supportati dalla comunità, coinvolgendo tassonomie vegetali e animali, nomi comuni, anatomia e terminologia di descrizione di alimenti per umani e animali domestici, riducendo così il costo e il carico di lavoro per chiunque implementi tale standard. Il vocabolario di FoodOn deriva dalla trasformazione di LanguaL¹ in un vocabolario OWL che fornisce interoperabilità al sistema, controlla di qualità e intelligenza basata sul software.

1.4 Obiettivi

I vari obiettivi che questo progetto si pone si possono già desumere dalle sezioni precedenti e possono essere riassunte e raggruppate in:

- promuovere un vocabolario di descrizione alimentare che la ricerca e le applicazioni di consumo e industriali possono riutilizzare;
- creare uno standard che coinvolge tassonomie vegetali e animali, nomi comuni, anatomia e terminologia di descrizione di alimenti per umani e animali domestici;
- creare un'ontologia che permetta di armonizzare i dati che abbracciano molteplici settori, dalla sicurezza alimentare alla salute e convenienza dei consumatori;
- interoperabilità con altre ontologie;
- indipendenza dalla lingua e dalla cultura.

¹LanguaL, che sta per “**L**angua **a**Limentaria”, è un dizionario di indicizzazione alimentare avviato alla fine degli anni '70 dal Center for Food Safety and Applied Nutrition (CFSAN) della Food and Drug Administration degli Stati Uniti grazie alla cooperazione di specialisti in tecnologie alimentari, scienze dell'informazione e nutrizione.

Capitolo 2

FoodOn

2.1 Ontologia

La necessità di rappresentare la conoscenza del cibo è fondamentale per molte attività umane tra cui agricoltura, medicina, ispezione della sicurezza alimentare, modelli di acquisto e sviluppo sostenibile. FoodOn è una risorsa ontologica completa e open-source composta da sfaccettature gerarchiche di termini che comprendono ingredienti di base delle materie prime alimentari, termini legati al processo per l'imballaggio, la cottura e la conservazione e una varietà di schemi dei tipi di prodotto in base ai quali i prodotti alimentari possono essere classificati. Ad ora FoodOn è fornito sotto forma di Web Ontology Language (OWL) tramite repository GitHub del progetto stesso: github.com/FoodOntology/foodon

L'ultima versione può anche essere esplorata tramite servizi di ricerca ontologica come:

- BioPortal, bioportal.bioontology.org
- European Bioinformatics Institute Ontology Lookup Service (EMBL-EBI), www.ebi.ac.uk/ols/
- Ontobee, ontobee.org
- AgroPortal, agroportal.lirmm.fr

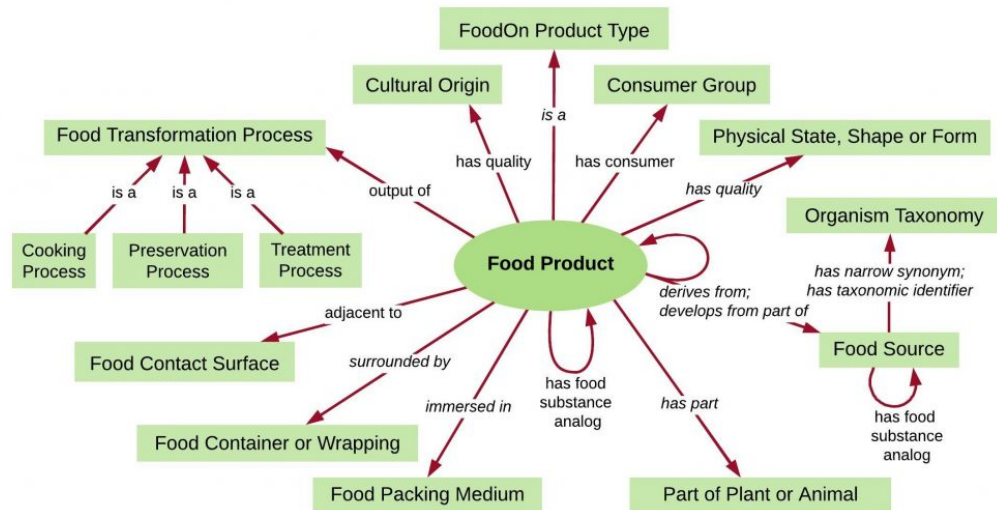


Figura 2.1: **Diagramma di “Food Product”** - Lo schema dei prodotti alimentari di FoodOn deriva principalmente dalle sfaccettature della descrizione degli alimenti di LanguaL con l’aggiunta di relazioni ontologiche tra un prodotto alimentare e le relative qualità descrittive, componenti e processi.

2.1.1 FoodOn e LanguaL

Come già accennato, gran parte del vocabolario principale di FoodOn deriva da LanguaL, un dizionario di indicizzazione degli alimenti maturo e popolare. In questa sezione si affronteranno le varie similitudini e differenze tra il nuovo e il vecchio progetto.

Similitudini

Come LanguaL, FoodOn fa la stessa distinzione tra le piante di origine alimentare e gli organismi animali che si trovano nella Facet B di LanguaL e gli schemi di categorizzazione degli alimenti delle varie aziende alimentari sempre all’interno della stessa Facet i quali sono tutti copiati termine per termine in FoodOn. Sotto il ramo “*food product type*” c’è una nuova gerarchia “*foodon food product*” che è una copia molto approssimativa del

ramo relativo ai prodotti alimentari del FDA Code of Federal Regulations (CFR), ma riorganizzata per essere più ontologicamente “pura”. La maggior parte degli articoli di “*foodon food product*” sono collegati direttamente o tramite eredità attraverso un riferimento “*member of*” ad una categoria CFR. L’intera Facet B di LanguaL è duplicata sotto il ramo “*food source*” di FoodOn ed un riferimento a un organismo nella Facet B non richiede che tale organismo svolga un ruolo alimentare. Di conseguenza si può fare un riferimento, ad esempio, a *Bovini* attraverso assiomi che possono essere completamente dedicati alle indagini epidemiologiche sulle malattie degli zoccoli senza implicare alcuna conversazione sui *Bovini* come cibo.

Differenze

FoodOn differisce da LanguaL su una serie di fronti. In primo luogo in LanguaL esistono casi in cui sono presenti termini del tipo “[*facet X*] not used” o “[*facet X term*] not known” come, ad esempio, “*Packing medium not used*”. In FoodOn si preferisce non menzionare aspetti se non sono pertinenti, ciò significa che se non si sa se un alimento può essere descritto o meno con un particolare termine, ma si sa che la categoria genitore è pertinente, allora si fa riferimento alla categoria genitore. Per descrivere un alimento in LanguaL si avrà, ad esempio, “*Cut of meat not known*”, invece in FoodOn “*Cut of meat*”. Ciò permette a un ragionatore di sapere che l’entità descritta è un *taglio di carne*, il che consente che una qualsiasi sottoclasse di *taglio di carne* sia pertinente.

In più LanguaL, per indicare l’ingrediente principale di qualsiasi alimento indicizzato, utilizza “*food source*” e per indicare gli ingredienti secondari “*treatment applied > ingredient added*”. FoodOn, grazie alle ontologie OWL che permettono di usare le proprietà degli oggetti per descrivere gli ingredienti primari e secondari, può utilizzare relazioni del tipo “*has ingredient*” e “*has substance added*”.

Invece, per regolamentare LanguaL, FoodOn ha conservato una serie di contenuti aggiuntivi del cibo. Tuttavia le relazioni “*has substance added*” di un prodotto alimentare potrebbero contenere delle percentuali o delle proporzioni, il che avvicina gli alimenti composti a una ricetta o a una descrizione dell’etichetta del prodotto.

Infine, la maggior parte delle voci dell'elenco degli additivi alimentari di LanguaL sono state convertite in voci ChEBI¹ equivalenti.

2.1.2 Gerarchie

Fonti alimentari

LanguaL utilizza la sfaccettatura “*food source*”, avente circa 3400 termini, per descrivere ogni singola fonte di cibo vegetale, animale o chimica da cui deriva il prodotto alimentare o il suo ingrediente principale. FoodOn rispecchia molto questa sfaccettatura di LanguaL introducendo, però, gruppi intermedi come “*stem or spear vegetable*” e spostando i prodotti chimici in una classe “*food component class*” per separarli dai riferimenti all'intero organismo. Gli organismi di origine alimentare di LanguaL hanno spesso delle specie associate e/o degli identificatori tassonomici di livello superiore provenienti dall'Integrated Taxonomic Information System (ITIS). Queste informazioni tassonomiche sulle specie di LanguaL sono conservate da FoodOn come annotazioni incrociate del database. Inoltre, se il riferimento ITIS di un termine FoodOn può essere mappato ad un elemento di una risorsa NCBITaxon², FoodOn utilizza una relazione di tipo “*has taxonomic identifier*” per andare a collegare le due facilitando così l'accesso a informazioni tassonomiche NCBI e altre informazioni collegate.

Parti vegetali o animali

In un linguaggio colloquiale, alcuni termini possono anche riferirsi ad un intero organismo commestibile, come *acciuga* o *cavalletta*, mentre altri si riferiscono solo ad una parte di un organismo. Ad esempio: *bacca* è una parte e *cespuglio* è l'intero; *mela* è una parte e *albero* è l'intero. Inoltre quelle parti di un organismo non sono sempre presenti o commestibili. La sfaccettatura “*part of plant or animal*” di LanguaL è definita come parte anatomica della pianta o animale da cui deriva il prodotto alimentare o il suo ingrediente principale. FoodOn riprende la maggior parte di questi

¹Chemical Entities of Biological Interest (ChEBI) è un database e ontologia riguardante entità molecolari focalizzate su composti chimici “piccoli”, che fa parte di Open Biomedical Ontologies.

²L'ontologia NCBITaxon è una traduzione automatica del database tassonomico dell'NCBI in OBO/OWL.

descrittori, ma riutilizza identificatori di termini ontologici Uberon³ e PO⁴. Ciò comporta che in FoodOn un prodotto alimentare come *apple* può essere definito come:

‘apple (whole) food product’ SubClassOf: ‘pome fruit’ and ‘develops from part of’ some ‘apple tree as food source’.

Sostanze analoghe e allergeni

FoodOn dà la possibilità di collegare due prodotti o ingredienti di origine alimentare attraverso la relazione “*has food substance analog*” permettendo così di indicare i suoi possibili sostituti per l’uso in analisi e in applicazioni sensibili agli allergeni e/o ad altri vincoli dietetici. Ciò consente, ad esempio, di associare la vaniglia naturale a quella sintetica. Tale relazione simmetrica non ipotizza quale parte stia imitando l’altra, né la qualità della sostituzione o il rapporto più appropriato. Inoltre, i termini relativi alle fonti alimentari di FoodOn relativi alle malattie allergiche da ipersensibilità vengono citati all’interno di Disease Ontology (DO).

2.1.3 Riutilizzo dell’ontologia

Come già accennato nel primo capitolo, uno degli obiettivi di FoodOn è quello di diventare uno standard ben documentato, curato e stabile. Grazie alla sua appartenenza alla OBO Foundry consente l’accesso a ontologie che coprono domini come: demografia dei consumi, composizione agricola e antimicrobici, tassonomie e fenotipi delle malattie. (*Figura 2.2*)

FoodOn non funge solo da hub generalista che si interfaccia con ontologie di dominio più specialistiche che coinvolgono la modellazione delle scienze alimentari, ma mira a coprire i prodotti alimentari e le ampie fasi di trasformazione degli alimenti. Ciò lo si può ben vedere dalla *figura 2.3*.

Il riutilizzo dei termini fra le varie ontologie consente di condividere lo sforzo e di fornire un vocabolario standardizzato. Per questo OBOFoundry incoraggia, con alcune eccezioni, ogni ontologia a riutilizzare i termini di

³Uberon è un’ontologia comparativa dell’anatomia che rappresenta una varietà di strutture anatomiche presenti negli animali come polmoni, muscoli, ossa, piume e pinne.

⁴Plant ontology (PO) è una raccolta di ontologie sviluppata da Plant Ontology Consortium. Queste ontologie descrivono le strutture anatomiche e le fasi di crescita e sviluppo attraverso Viridiplantae.



Figura 2.2: FoodOn riutilizza i termini di ontologie di OBOFoundry.org

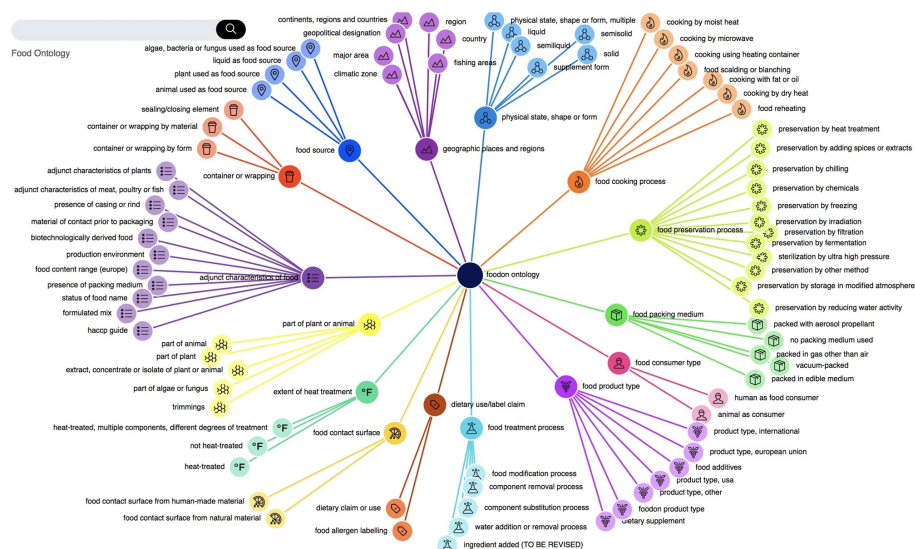


Figura 2.3: Visualizzazione di uno schema ad albero di FoodOn.

altri dove applicabile. Nel caso di FoodOn, sono stati sostituiti circa 600 termini di sostanze chimiche di LanguaL con identificatori chimici ontologici di ChEBI. Le eccezioni accennate precedentemente riguardano al fatto che le ontologie di OBO Foundry devono rispettare una determinata struttura tecnica generale che riguarda la le migliori pratiche per la cura delle versioni dei file delle ontologie, gli URL permanenti per i termini e gli schemi utilizzati per annotare le sostituzioni di termini deprecati in modo che il contenuto del database possa essere aggiornato senza problemi.

Riprendendo la *figura 2.1* c'è da chiarire che molte delle relazioni provengono dall'OBO Foundry's Relation Ontology il che porta restrizioni al dominio delle relazioni OWL e al codice stesso. In FoodOn queste relazioni sono combinate quelle di Basic Formal Ontology. Ciò consente a un ragionatore di applicare un riferimento adeguato a processi, qualità, ruoli ed entità informative.

Esempio:

'hen (food source)' : 'chicken (food source)' and ('has quality' some 'female organism') and ('has quality' some 'adult organism')

ciò porta ad un contraddizione perché la relazione *'has quality'* in BFO consente solo qualità nel suo range, ovvero sul proprio lato destro, e sia *'female organism'* che *'adult organism'* sono caratteristiche di un'entità materiale essendo un tipo di *'independent continuant'* il quale è disgiunto rispetto alle qualità. Più semplicemente, le qualità sono caratteristiche delle entità materiali ma non possono essere esse stesse entità materiali. Quindi la definizione può essere formulata nel seguente modo senza generare contraddizioni:

'hen (food source)' : 'chicken (food source)' and 'female organism' and 'adult organism'

Per l'incorporazione di nuovi contenuti ontologici sarebbe necessaria una certa padronanza della tecnologia Web Semantico, inclusa la conoscenza di OWL e del linguaggio di query SPARQL. Tuttavia viene incoraggiato il riutilizzo dell'ontologia anche per chi non possiede questa formazione completa fornendo fogli di calcoli o modelli di database personalizzabili tramite portali web. Inoltre, Hsiao Lab sta sviluppando uno strumento che trasforma i contenuti ontologici di markup in standard, forniti sia tramite una form web che in versione tabulare o json per l'implementazione nei sistemi di gestione e scambio dati.

2.2 Struttura

Nella sezione precedente si è accennato al fatto che FoodOn è compatibile con Basic Formal Ontology (BFO), il che significa che tutte le classi fornite da FoodOn sono organizzate in classi BFO. In realtà FoodOn include BFO, quindi quando si esplora l'ontologia in maniera top-down è un po' più complicato notare dove finisce BFO e dove iniziano i termini di FoodOn. Di seguito saranno mostrate le principali classi di FoodOn.

“**food material**” è un tipo di “**environmental material**” che è stato posizionato sotto “**fiat object part**” perché il confine tra materiali ambientali adiacenti può essere difficile da definire fisicamente e questo si ripercuote anche su parti commestibili e non di entità alimentari. (*Figura 2.4*)

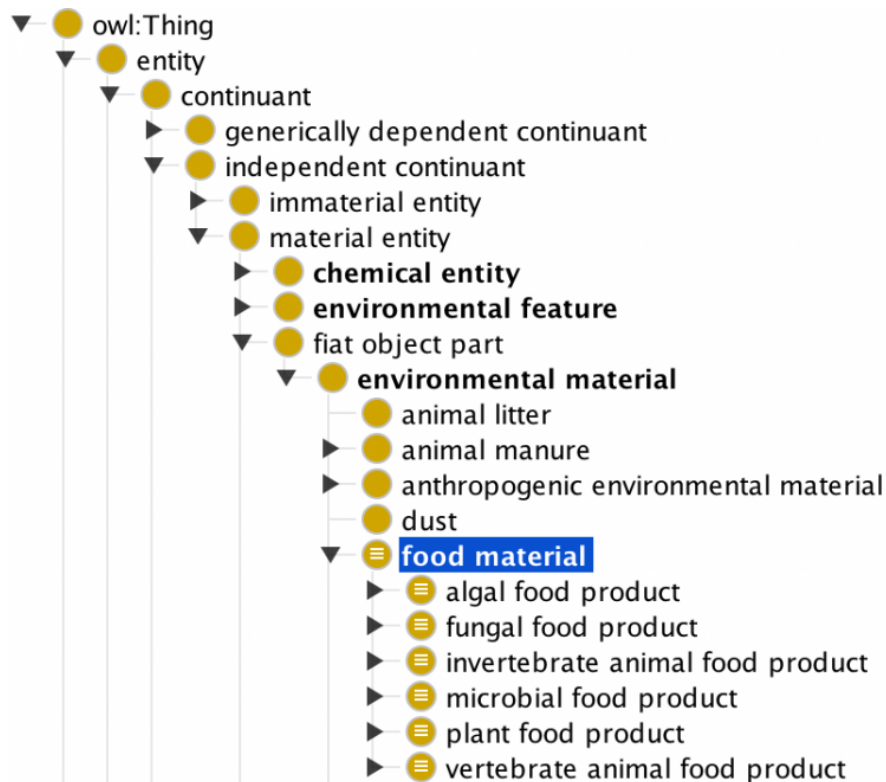


Figura 2.4: food material

Altre classi chiave di FoodOn come quelle legate a contenitori di cibo, involucri alimentari, mezzi di imballaggio, struttura della pianta, gruppi di consumatori e sostanze chimiche costituenti si trovano sotto la classe “**material entity**”. (Figura 2.5)

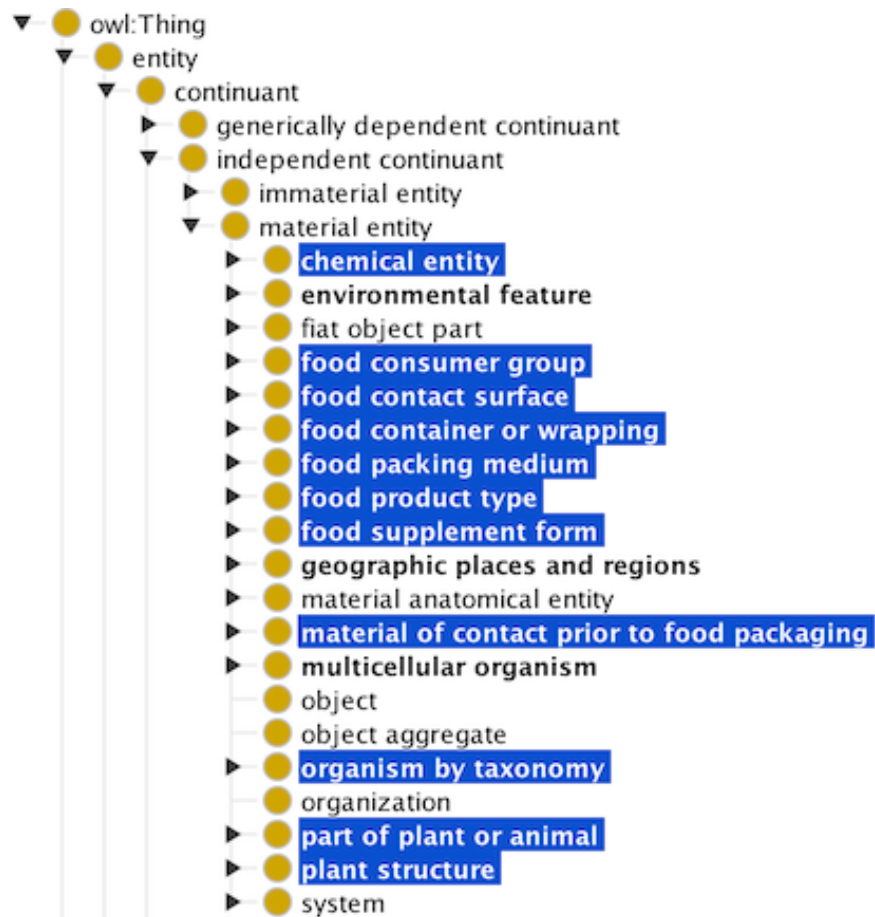


Figura 2.5: material entity

La gerarchia “**food product type**” contiene la principale gerarchia di FoodOn: “**foodon product type**”. “*food product type*” contiene oltre 9600 prodotti alimentari, nonché schemi di classificazione alimentare internazionale, europea e americana di livello superiore che verranno infine associati alle categorie di “*foodon product type*”. (Figura 2.6)

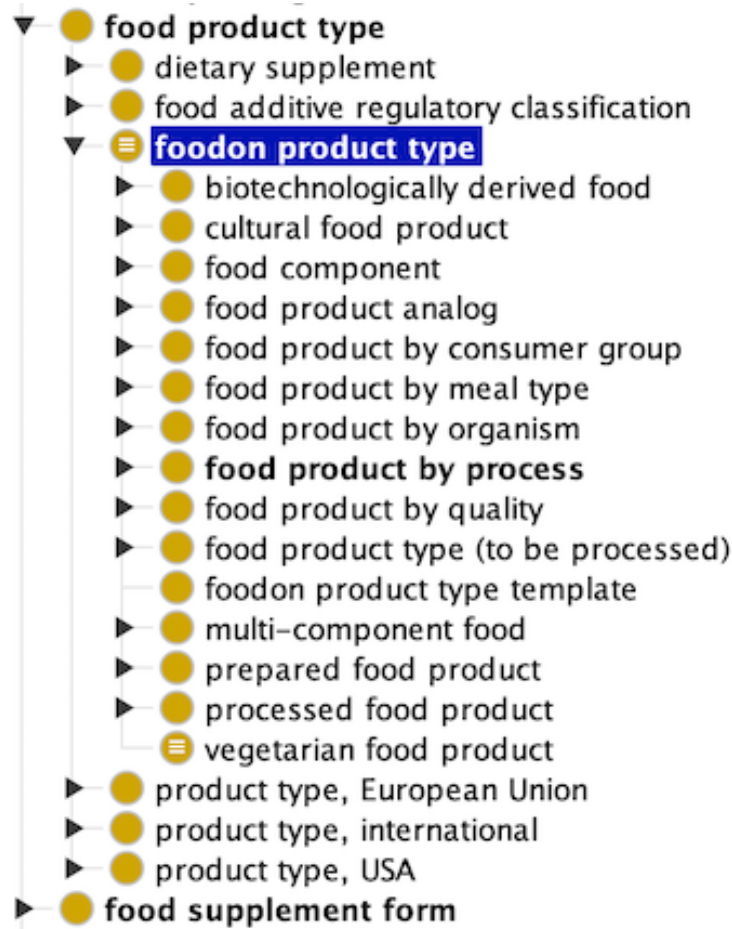


Figura 2.6: food product type

La classe “**food transformation process**” è posizionata sotto la classe BFO “**process**”. Alcuni processi di FoodOn possono essere multiuso, ad esempio, il congelamento può essere per la conservazione degli alimenti o per produrre un effetto come nel gelato. Altri sono specifici per la conservazione o altri obiettivi del processo. (*Figura 2.7*)

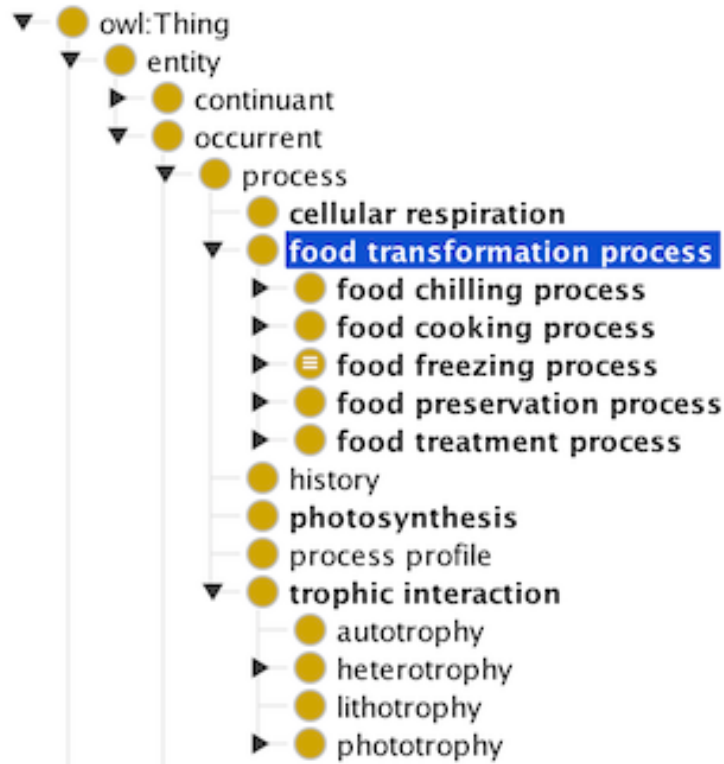


Figura 2.7: food transformation process

Sotto la classe “**quality**” si trovano le qualità osservabili del cibo come, ad esempio, la consistenza della carne del pesce o il colore della polpa o della buccia di un frutto. (*Figur 2.8*)

NB: C’è anche una classe BFO “quality”: le due non sono ancora unificate ontologicamente.

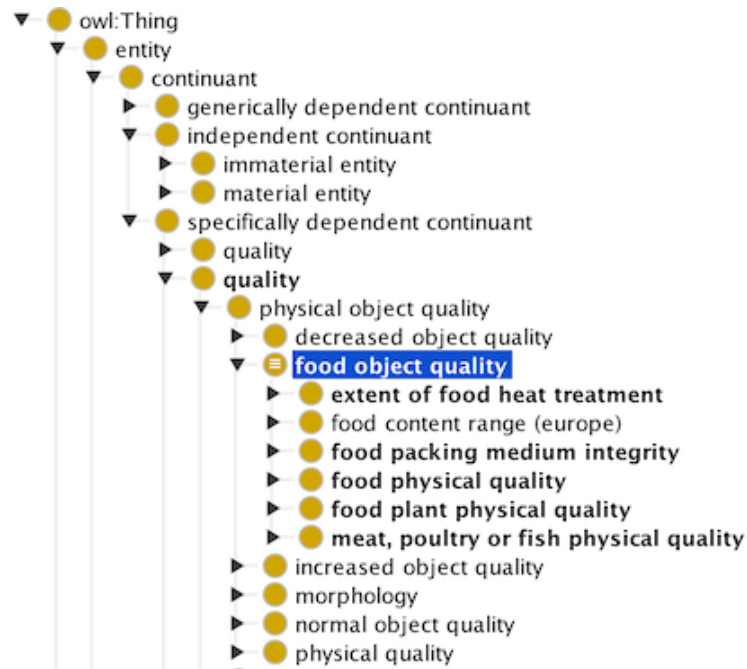


Figura 2.8: quality

Alcune categorie di informazioni che riguardano gli alimenti e che di solito non sono viste direttamente come qualità di un prodotto alimentare, ma piuttosto sono informazioni più legate a fatti legislativi o affermazioni del produttore sull'articolo, di trovano in “**information content entity**”. (Figur 2.9)

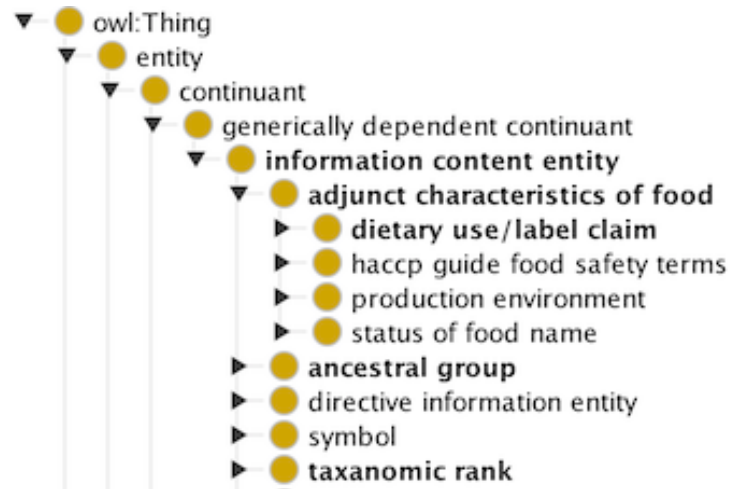


Figura 2.9: information content entity

2.3 Relazioni

Per quanto riguarda alle relazioni FoodOn utilizza:

- “**has part**”, utilizzato tra entità dello stesso tipo, ad esempio:
 - tra entità anatomiche: *“albumen is part of an egg”*
 - tra parti prescrittive: *“a pie has a crust and a filling”*
- “**has substance added**”, proprietà di un oggetto che può essere applicata a una qualsiasi sostanza chimica, fisica, alimentare o altro. Si differenzia da “has part”, perché la sostanza aggiunta potrebbe non essere individuabile nella miscela in un secondo momento.
- “**has ingredient**”, sotto-proprietà di “has substance added” pertinente solo a un dominio e intervallo alimentare. È usato per riferirsi a una sostanza alimentare aggiunta a un prodotto alimentare.
- X “**member of**” Y, utilizzato per associare un prodotto alimentare a una categoria di tipo prodotto che non rientra nel ramo “foodon product type”.
- X “**develops from part of**” Y può essere usato quando X è una struttura cellulare che subisce una crescita biologica. La relazione non implica l’eliminazione di Y, infatti molte piante generano cibo da raccogliere durante la stagione di crescita o comunque per diversi anni. In FoodOn, questo viene applicato principalmente a frutta, semi e verdure come le zucchine raccolte dalle piante.
- Y “**produced by**” X viene utilizzato quando Y è una sostanza che non subisce una crescita biologica, come, ad esempio, il latte.
- X “**derives from**” Y viene utilizzato quando non sono ancora disponibili o curati rapporti più specifici. Spesso funge da scorciatoia per un modello di processo più elaborato, se presente.
- X “**has food substance analog**” Y consente di collegare due prodotti quando hanno proprietà fenotipiche o ruoli di processo simili.

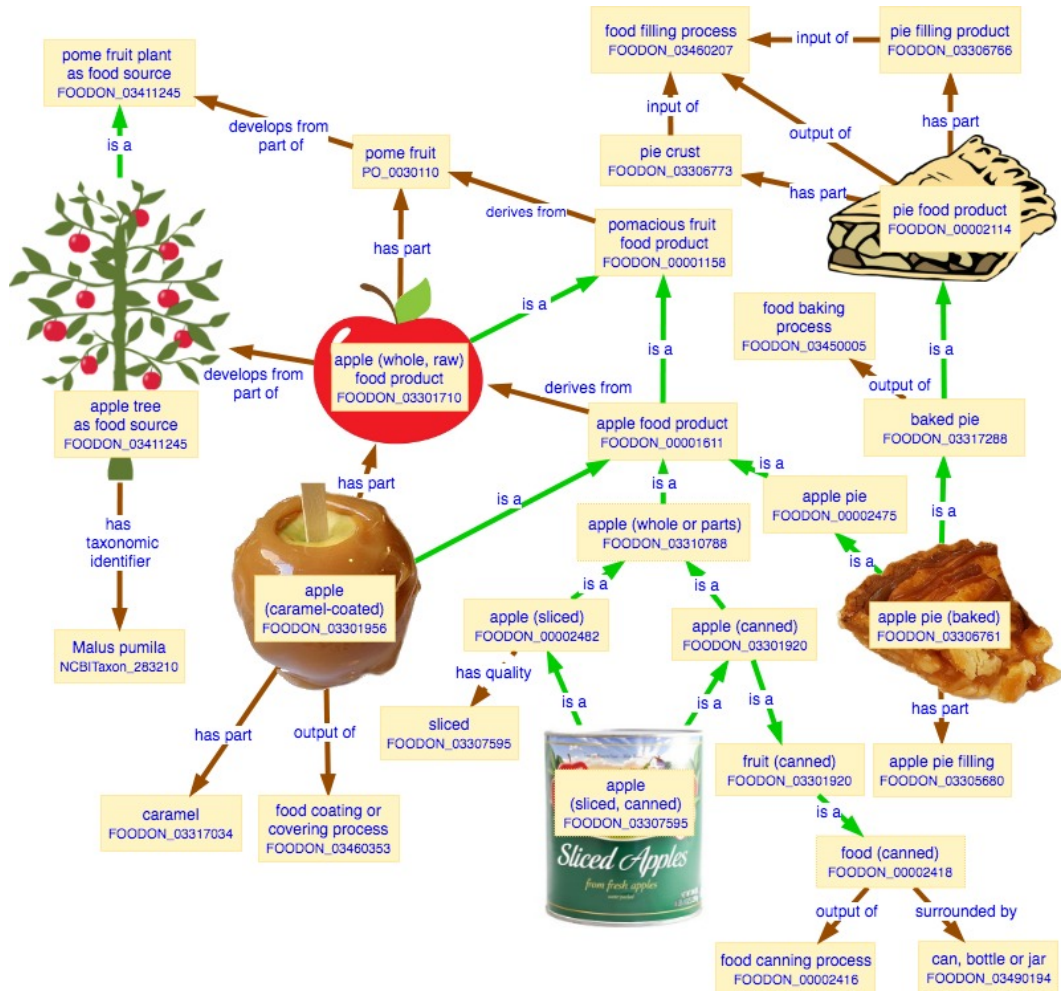


Figura 2.10: Diagramma che esplora la fonte di cibo “melo” nelle relazioni tra prodotti alimentari. Si possono notare tre diversi tipi di prodotti correlati alle mele mostrati in relazione a un organismo di origine alimentare, il melo, e vari processi di preparazione e confezionamento degli alimenti.

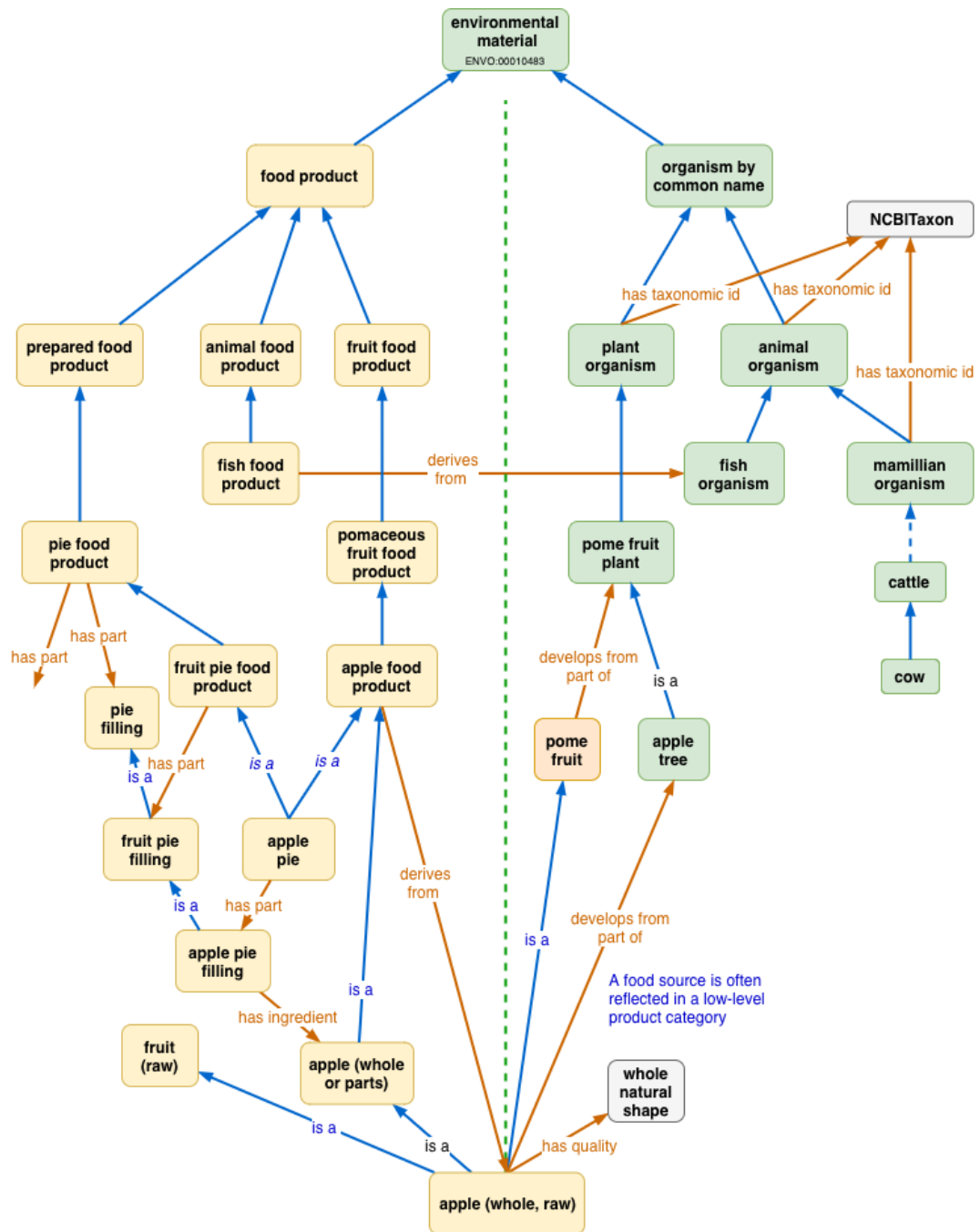


Figura 2.11: Schema mostra alcuni altri tipi di prodotti alimentari e le varie relazioni che FoodOn impiega per collegarli a organismi di origine alimentare.

Capitolo 3

Conclusioni

