

Link utili: <http://geneontology.org/docs/literature/>

Data la mole di dati in costante crescita che ruota attorno alla biologia molecolare, le ontologie sono emerse come strumento computazionale essenziale per aiutare l'organizzazione, la descrizione e l'analisi dei dati.

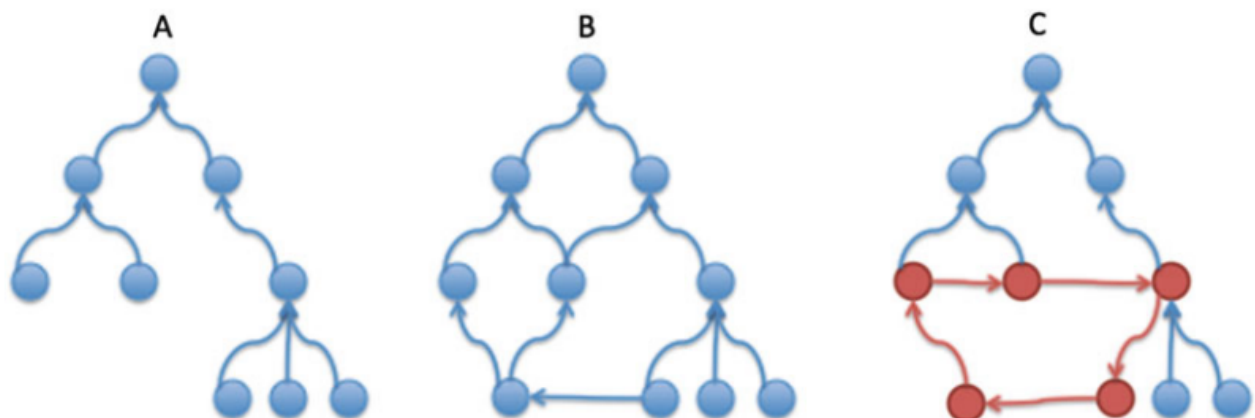
Le ontologie descrivono e classificano le entità di interesse in uno specifico dominio scientifico in maniera accessibile in modo tale da poter sviluppare algoritmi e strumenti su di esse. La tecnologia alla base delle ontologie ha le sue radici nell'intelligenza artificiale basata sulla logica, consentendo una sofisticata inferenza automatizzata e il rilevamento degli errori.

Le ontologie sono costituite da diversi elementi distinti, fra cui classi, metadati, relazioni formali e assiomi.

Classi: la classe è l'unità base di una ontologia e rappresenta un tipo di una cosa in un dominio di interesse. Esempio: l'acido carbossilico, il cuore, il melanoma e l'apoptosi. Solitamente le classi sono associate ad un identificatore univoco, per esempio (rispettivamente) CHEBI:33575, FMA:7088, DOID:1909 e GO:0006915. Tali identificatori sono privi di semantica, non contengono riferimento al nome o alla definizione della classe, questo per promuovere stabilità anche nel momento in cui la conoscenza scientifica e la relativa rappresentazione evolvono.

Metadati: I metadati associati alle classi possono includere qualsiasi identificativo secondario associato e indicatori per indicare se la classe è stata contrassegnata come obsoleta. Possono anche includere uno o più sinonimi o riferimenti incrociati a quella classe in database alternativi e risorse web.

Relazioni: le classi sono disposte in una gerarchia che va dal generale allo specifico. Nonostante la classificazione gerarchica, la maggior parte delle ontologie non sono semplici alberi, piuttosto sono strutturate come grafi aciclici diretti. Questo perché è possibile che le classi abbiano più genitori nella gerarchia di classificazione e inoltre le ontologie comprendono altri tipi di relazioni tra entità diverse dalla classificazione gerarchica. Tutte le relazioni sono dirette e i redattori dell'ontologia devono fare attenzione a garantire che la struttura complessiva non contenga cicli.



(a) A simple hierarchical tree, (b) a directed, acyclic graph, (c) a graph that contains a cycle, indicated in red

Un tipo di relazione comune usato in più ontologie è **part_of** o **has_part**, che rappresenta la composizione o la costituzione. La specifica di un tipo di relazione in un'ontologia include un identificatore univoco, un nome e una gerarchia di classificazione. Gli stessi metadati che sono associati alle classi possono essere associati anche ai tipi di relazione.

Formati: In genere le ontologie vengono memorizzate in file conformi a un formato specifico, anche se esistono eccezioni che vengono memorizzate in infrastrutture personalizzate. Le ontologie possono essere rappresentate in diversi linguaggi ontologici e storicamente si è assistito a un'evoluzione delle capacità dei linguaggi ontologici verso una maggiore espressività e complessità logica, che è stata rispecchiata dai proressi della capacità computazionale e degli strumenti. Le ontologie biologiche come GO sono state storicamente rappresentate nel linguaggio OBO (Open Biomedical Ontologies), progettato specificamente per la struttura e il contenuto dei metadati associati alle bio-ontologie, ma col tempo ci si è spostati verso lo standard del web semantico, il Web Ontology Language (OWL).

Assiomi: Nei linguaggi logic-based come OWL, le affermazioni nelle ontologie hanno un significato logico definite all'interno di una teoria logica basati su insiemi. Le classi hanno istanze come membri, e gli assiomi logici definiscono vincoli sulle definizioni di classe che si applicano a tutti i membri della classe. I linguaggi logici alla base della tecnologia ontologica sono chiamati collettivamente Logiche di Descrizione. Alcuni degli assiomi logici disponibili nel linguaggio OWL: quantificazione, cardinalità, connettivi logici e negazione, disgiunzione ed equivalenza di classi, ecc...

Funzione: la nozione di funzione in biologia ha ricevuto una grande attenzione. Esistono due scuole di pensiero su come dovrebbero essere definite le funzioni: "funzione di ruolo causale" e "funzione di effetto selezionato". La funzione di ruolo causale è stata proposta per la prima volta da Cummins [1] e si concentra sulla descrizione della funzione in termini di come una parte contribuisce a una capacità complessiva del sistema che la contiene. In questa formulazione, la funzione di un'entità è relativa a un sistema a cui contribuisce. Per esempio l'affermazione "la funzione del cuore è quella di pompare sangue" ha significato solo nel contesto della capacità del sistema circolatorio più ampio di fornire nutrienti e rimuovere prodotti di scarto dai tessuti corporei. Tuttavia una delle principali obiezioni alla definizione di ruolo causale è che non esiste un modo sistematico di identificare quale debba essere il sistema più ampio e la relativa capacità di tale sistema. La funzione di "effetto selezionato", invece, deriva dalla definizione "eziologica" di funzione proposta per la prima volta da Wright [2]. In questa formulazione, la funzione di un'entità è la risposta alla domanda sul perché l'entità esiste, ovvero equivale a chiedersi "per quale dei suoi effetti è stata selezionata durante l'evoluzione?". Un vantaggio evidente di questa formulazione è che incorpora esplicitamente considerazioni evolutive e richiede che una funzione derivi in ultima analisi dalla sua storia di selezione naturale. Ritornando all'esempio fatto in precedenza, un effetto del cuore è quello di produrre un suono, ma non sarebbe corretto dire che la funzione del cuore è quella di produrre un suono. La definizione di funzione con effetto selezionato distinguerebbe una funzione propria (pompare sangue) da un effetto accidentale (ad esempio produrre un suono) sulla base del fatto che la selezione naturale ha più probabilmente operato sull'effetto di pompare il sangue.

Funzioni in GeneOntology: per comprendere come le funzioni dei geni siano rappresentate in GO, sono necessarie alcune conoscenze di base di biologia molecolare:

- Un gene è una regione contigua di DNA che codifica le istruzioni per la produzione di una grande molecola da parte della cellula (o potenzialmente più macromolecole diverse).
- Una macromolecola è chiamata prodotto genico (in quanto viene prodotta secondo le istruzioni di un gene) e può essere di due tipi: una proteina o un RNA non codificante.
- Un prodotto genico può agire come una macchina molecolare, cioè può eseguire un'azione chimica che chiamiamo attività.
- Prodotto genici di geni diversi possono combinarsi in macchine molecolari più grandi, chiamati complessi macromolecolari.

Ogni concetto della Gene Ontology si riferisce all'attività di un prodotto o di un complesso genito, in quanto si tratta di entità che svolgono processi cellulari. Un gene codifica un prodotto genico, quindi può ovviamente essere considerato la fonte ultima di queste attività e processi. Ma in senso stretto, un gene non svolge un'attività in sé. Pertanto quando la Gene Ontology si riferisce alla "funzione genica" in realtà è un'abbreviazione di "funzione del prodotto genico".

La missione del Consorzio GO è quella di sviluppare un modello computazionale aggiornato e completo dei sistemi biologici, a partire dal livello molecolare fino ad arrivare a percorsi più ampi, sistemi cellulari e a livello di organismo.

La gene ontology fornisce una rappresentazione delle attuali conoscenze scientifiche sulle funzioni dei geni (più precisamente delle molecole proteiche e di RNA non codificanti prodotte dai geni) di molti organismi, dall'essere umano fino ad arrivare ai batteri. È ampiamente utilizzata a supporto della ricerca scientifica e conta decine di migliaia di citazioni nelle pubblicazioni. Capire la funzione dei geni è uno degli obiettivi principali della ricerca biomedica. Inoltre, le conoscenze ottenute su un organismo sono spesso applicabili ad altri organismi, soprattutto se questi condividono i geni in questione perché li hanno ereditati entrambi da un antenato comune.

La gene ontology come consorzio è nata nel 1998 quando i ricercatori che studiavano il genoma di 3 organismi modello - *Drosophila melanogaster* (moscerino della frutta), *Mus musculus* (topo) e *Saccharomyces cerevisiae* (lievito di birra) - hanno deciso di lavorare in collaborazione su uno schema di classificazione comune per la funzione dei geni. Oggi la gene ontology vanta migliaia di organismi rappresentati.

L'ontologia consente, in modo flessibile e dinamico, di fornire descrizioni comparabili di sequenze di geni e proteine omologhe in tutto lo spettro filogenetico.

È collegata a molte altre ontologie biomediche e costituisce una base per la ricerca che applica l'informatica alla biologia e alla medicina.