An ontology is essentially a representation of relationships between observable attributes/features. Gene ontology (GO) is a repository of gene/gene product associated attributes and relationships between these attributes. A common record of the relationships and attributes describing a given gene (or gene product) allows us to resolve semantic differences between meaning and naming conventions that may otherwise arise between genomic data/annotations that have been generated at different sites. GO is composed of three ontologies (domains), which are species independent. These three ontologies respectively record the biological processes, cellular components, and molecular functions associated with a gene/gene product. An attribute can hence be classified under either of these 3 domains. The attribute itself can be a biological term (ex. Cell cycle), a definition, a unique alphanumeric identifier, or just a (string of) word(s).

The attributes are represented on a directed acyclic graph, and the edges between the nodes can be considered equivalent to the relationships (like is\_a, regulates,positively\_regulates) between attributes. Generally, the edges only link nodes that are in the same domain – but attributes from different domains can be associated via an edge as well (ex. Cell cycle [biological processes] and base excision repair [molecular functions]). These attributes (and relationships between attributes) are associated to genes/gene products via an ‘evidence code ontology’, which computationally infers and records the type of reference (ex. Sequence similarity, traceable author statement) used to make an annotation about a gene (product). GO ontologies are generally recorded in the Open Biomedical Ontology (OBO) flat file structure. The data file recording this giant ontology can also be downloaded in various formats (OWL-RDF/ XML, OBO). The extended set of cross-ontology relationships and external ontologies (like ChEBI, cell ontology, Uberon) are only available in the OWL format, however. OWL Web Ontology Language has increased machine interpretability of relationships and terminology as compared to XML and RDF formats. This facilitates data input to machine learning algorithms, and also increases ease of representation of machine interpretable content on the Web. OBO is an ontology language similar to OWL, but records unidirectional relations (it does not record the inverse of a relationship, unlike OWL). The relations are hence solely stored in a ‘wider’ to ‘narrower’ perspective, based on the order they occur in the OBO flat file.

Gene Ontology falls under an umbrella initiative of developing biological ontologies – Open Biomedical Ontologies. Other popular ontologies are Sequence Ontology and Plant Ontology, AmiGO, Gene Expression Database (GXD), and Unified Medical Language System (UMLS). Among other things, these ontologies are actively being used in several bioinformatics applications like expression profiling and biological knowledge graph construction (especially UMLS).

Reference:

<http://geneontology.org/page/documentation>

<http://www.w3.org/TR/owl-features/>

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2735951/>

Bard, J. B. L. & Rhee, S. Y. Ontologies in biology: design, applications and future challenges. Nature Reviews Genetics **5**, 213-222 (2004) doi:10.1038/nrg1295.