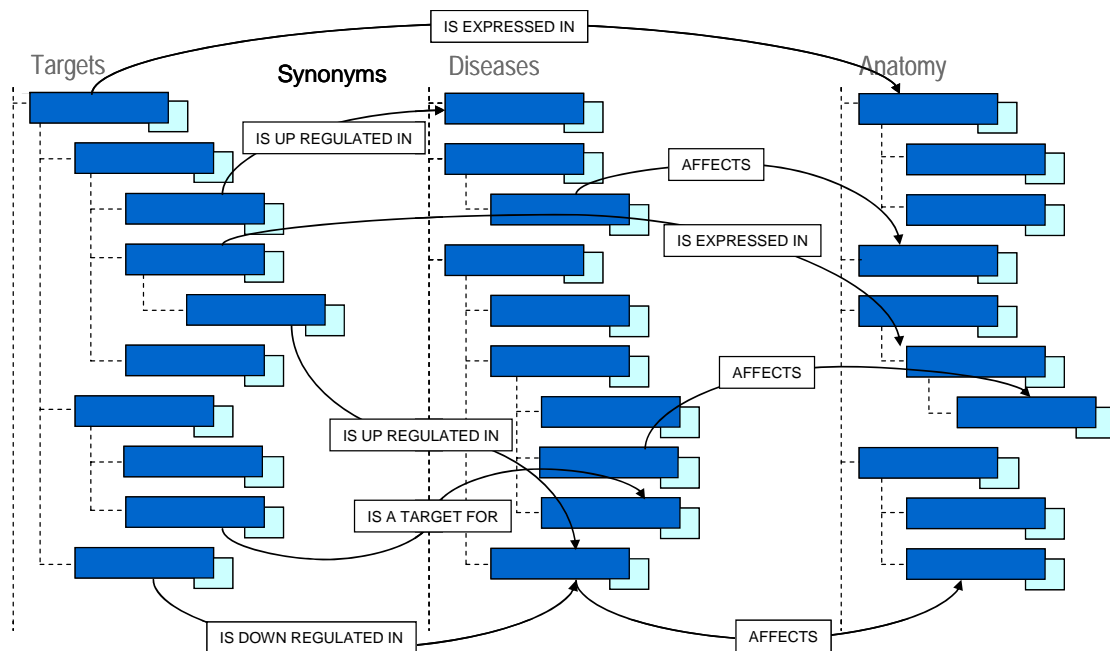


## Amgen Project Description

For Amgen, an increasingly serious business problem is that scientists are unable to find and integrate all of the data involved in answering complex scientific questions. These data sources are principally unstructured (journal articles, books, internal reports, presentations, etc.), and any approach to solving the problems must take into account that newly evolving scientific information includes significant overlaps and can even contradict each other. Such contradictions provide essential information in pointing the way forward for researchers.

To organize all this information, knowledge management workers are increasingly turning to ontologies. Ontologies enable integration and communication between systems and scientists by establishing a shared understanding. One kind of ontology handles structured data that is largely self-consistent. Our needs are for a model that explicitly exploits the rich and complex nature of linguistic terms to manage the diffuse, overlapping, and sometimes contradictory assertions about connections between these terms. A subset of a possible literature-based ontology is presented graphically below, illustrating an ontology as interconnected taxonomies.



Amgen has developed a large complex biological ontology resulting from massive literature searches around a specific set of topics of interest. Exploring linguistic ontologies is difficult, and the existing building blocks for access and navigation are very low-level and *ad hoc*. Our scientific clients need a scalable, systematic set of tools with which to interrogate and analyze subsets of the extracted literature, and for this we need a more abstract and general set of building blocks.

The objectives of this project will be:

- Develop appropriate (recursive) algebraic datatypes allowing corresponding recursive processing of ontological information
- Apply a functional programming approach to link and query linguistic concepts systematically
- Realize this approach as an appropriately scalable set of combinators.

Deliverables

- A prototype library of combinators in the functional programming language Haskell
- An evaluation of the combinatory library on our existing ontology, including an understanding of the sensitivity of the constructed ontology to parameters of the literature search
- A technical report describing the project's accomplishments in detail

References:

- R. Bird, *Introduction to Functional Programming using Haskell*, Prentice Hall, 1998
- Wikipedia article on (the computer science concept of) ontologies: [http://en.wikipedia.org/wiki/Ontology\\_%28computer\\_science%29](http://en.wikipedia.org/wiki/Ontology_%28computer_science%29)
- Wikipedia article on combinators: [http://en.wikipedia.org/wiki/Combinatory\\_logic](http://en.wikipedia.org/wiki/Combinatory_logic)