When working with ontologies, a major concern in the semantic web, there is a hurdle to overcome with the decentralized definition of ontologies. Each individual or organization can define ontologies on their own with no oversight by a central body. This means that the same concept that is defined in one ontology can be defined again in another ontology and may even be defined different. Another case is that two different ontologies use the same terminology for differing concepts. These two scenarios are defined by Gracia and Mena as:

**Semantic Ambiguity:** many intended meanings are associated with the same word.

**Semantic Redundancy:** many semantic descriptions are available to represent the same intended meaning.

An example of this problem is trying to determine the meaning of the word “apple” in a query such as “Give me a list, ordered by calories, of recipes containing apple”. In the example query the term apple is likely apparent to a human reader, but to a device trying to determine the semantics of the word it can be difficult. If “apple” is run through an ontology it might come back with three results, apple the fruit, apple the tree, and apple the corporation. With these three results the query processor needs to determine what concept represented by apple does the user want to see the results of.

In order to find the correct concept, a search may cross many ontologies where each may have a definition of “apple” and its related concepts. This is where semantic ambiguity and redundancy start to have an effect. In order to filter the search down to likely candidates, word sense disambiguation techniques can be used.

There are first techniques to reduce semantic heterogeneity of the result set. There are three techniques proposed by Gracia and Mena for accomplishing semantic heterogeneity reduction. In ontology matching, a set of terms are compared to each other based on context and the relations that don’t meet a threshold are removed. The output of ontology matching then is fed into sense clustering.

In sense clustering the relatedness of two concepts that share keywords are used to determine if the two concepts should be integrated into a single concept. When two concepts are merged together, then the ontology matching algorithm needs to be run again to calculate the similarities of the new integrated concept.

After sense clustering comes sense disambiguation, in which the concept that the user is trying to query on is determined from the context of the query and the clusters created from sense clustering. This is done by taking the inputted query and breaking it down into keywords and context words to be utilized for the disambiguation then using then searching through related concepts in order to find the best fitting one. The first step is to break down the query into individual words and select a keyword and then compute the relatedness of each word to the selected keyword. This works on the assumption that the context with the highest relatedness value is the most likely candidate. The relatedness calculation can be done using something like Normalized Google Distance. With a selected context, the clusters from sense clustering are then weighted within the context to find the likely concept that should be used. The final selection can be done by calculating relatedness, number of word overlaps between concept and context, and determining frequency of appearance of the concept in the context overall.

Sense clustering

* Without a definitive standard to follow each implementer of a social media platform will create their own structure and definition set for representing the relationships and content in their social web.
* General problem of looking for “apple” and finding the fruit and the company
* Word Sense Disambiguation techniques are used to decide on the right meaning of a word
  + Meaning of a word is a “person” in social media
* Semantic Heterogeneity Reduction Techniques
  + Compute the degree of similarity and relatedness among different semantic descriptions
    - ontology matching - Redundancy
      * Finding similarities between terms and entities from different ontologies
    - sense clustering - Redundancy
      * Grouping and integrating representations into the same meaning
    - sense disambiguation - Ambiguity
      * The pick most appropriate meaning if multiple found
* Polysemous : Having multiple meanings
  + Polysemous. (n.d.). Retrieved March 6, 2016, from <http://www.merriam-webster.com/dictionary/polysemous>
* Semantic Measures allow for the comparison of how related two terms are, which can lead to the discovery of if the two terms are the same
  + This is how we can determine if Person A on social networking site A is the same as Person B on social networking site B
  + A lot of work and techniques in this area are designed around the search of words and phrases, not abstract concepts of individual
* Sense Clustering – Trying to find relationships across ontologies and determine which concepts within an ontology are
  + Initially groups all concepts that share a top level term
  + Computes similarity distances between same cluster elements
  + Use calculated distances to break apart clusters and merge neighboring clusters
* Sense Disambiguation – Selecting one of the clusters created/determined after Sense Clustering and selecting one as the final result for use.
  + Context Selection – Find things surrounding the concept in the search and in the clusters and remove context items that score below a certain threshold
  + Filtering – Use selected context to then filter clusters/senses based on their ability to be above the threshold of similarity
  + Selection – Finally run weights on the returned clusters and find out which one best fits the context based on overlap, frequency of use, and similarity

