CLAIMS

What is claimed is:

1. A method comprising:

determining, utilizing natural language understanding processing, first parts of speech for first words from first text data associated with a product, wherein the first text data is obtained from a first data source associated with a data processing system;

determining, utilizing the natural language understanding processing, second parts of speech for second words from second text data corresponding to one or more claims in a portfolio of patent documents, wherein the second text data is generated from a second data source;

determining a portion of the first words that correspond to a feature of the product;

determining, based at least partly on the first parts of speech, a first action performed with respect to the feature;

generating, based at least partly on the first action, first data representing a first linguistic structure for the feature, the first linguistic structure indicating one or more first relationships between second data representing the first action and third data representing one or more first nouns included in the first text data, the first linguistic structure including nodal links in a hierarchy;

determining, based at least partly on the second parts of speech, a second action included in the one or more claims;

generating, based at least partly on the second action, fourth data representing a second linguistic structure for the one or more claims, the second linguistic structure indicating one or more second relationships between fifth data representing the second action and sixth data representing one or more second nouns included in the one or more claims;

determining a similarity metric between the first linguistic structure and the second linguistic structure, the similarity metric indicating a similarity of an arrangement of nodes as between the first linguistic structure and the second linguistic structure; and

determining, based at least partly on the similarity metric, a portion of the portfolio of patent documents that corresponds to the product.

2. The method of claim 1, further comprising:

determining a first amount of similarity between the first action and the second action; and

wherein the similarity metric is determined at least partly based on the first amount of similarity.

3. The method of claims 1, further comprising:

determining an amount of similarity between a first word of a second node of the second level and a second word of a fifth node associated with the second level; and

wherein the similarity metric is determined at least partly based on the amount of similarity.

4. The method of any one of the preceding claims, wherein the one or more claims include elements, the action corresponds to an element of the elements, and the method further comprises determining that the element corresponds to the feature based at least partly on the similarity metric.

5. The method of claim 1, further comprising:

receiving input data corresponding to an amount of similarity between the portion of the portfolio and the product; and

modifying, based at least partly on the amount of similarity, the similarity metric.

6. A system comprising:

one or more processors; and

one or more computer-readable media storing instructions executable by the one or more processors, wherein the instructions, when executed by the one or more processors, cause the one or more processors to perform the method of any preceding claim.