**Universal Translator**

**Matter – Information Analogue**

Akin to matter, which in small, atomic and subatomic space, in the manner of combination, takes principal part of transition from small, identical units of few in number to the nexdt level of complexity from hundreds to millions to billions of combinations, . If we examine the principal network that in its individual pieces are alike tom subatomic particles, of which there are few distinct, but in combination have a great variety. Assuming a smalll number of stems, prefixes, and suffixes, as equivalent to subatomic particles, we can naturally see the analogue – an increase in complexity and variety of words.

**Thumbprint Theorem**

Every phrase has a unique finger print, recognizable in the form of a graph.

**Synopsis**

Running two language graphs against each other will allow for translation. One known language can be *almost* isomporhic to another,. The practice of graph matching, from the top level down to the individual atoms allows us to match fingerprints from langauge to language.

**Process is as follows:**

* The first step is to parse the plaintextx into a graph (how?)
* The second step is to determine the graph edit distance, which determines the shortest amount of operations to transform G to G’. A good method may be the Earth Movers Distance, Uses Caterpillar Trees and Sphereical (or Euclidean) Graph Embedding
  + Graph is first embedded into a tree
  + Tree is decomposed using caterpillar decomposition
  + Using the paths from caterpillar decomposition, embeds vertices into vector spcae using the spherical embedding method
  + With the vertices embeded in Euclidean Space, now we can apply earth movers distance algorithm to determine the cost of closing the holes with the piles.The little the cost, the higher the matching rate is.
  + Higher matching rate means the language subgraph is close to being synonymous.

**Caterpillar Decomposition**

Collection of edge-disjoint(sub) root-leaf paths.

**Caterpillar Decomposition continued**

The three paths between a and c, a and g, , a and m, are called level 1 paths and represent the first three paths in the caterpillar decompisition. If we remove these three paths from the tree, We are left with three edge dis-joint paths. These are the paths between e and f, I and j, k and l called level 2 paths. These represent the other three paths in the decomposition. If the level paths had left any connected components, the process would be repeated until all the edges have been removed.

**The union of these paths is called the caterpillar decomposition, denoted by . The total number of paths in the caterpillar decomposiion specifies the dimensionality of the geometric space into which the graph is embedded.**

Caterpillar Decomposition Explained

The tree is traversed from the root to the leaf to form a path. Upon reaching a junction, the larger value is taken and the traversal continues.

This is completed until all branches from the root have been traversed and added to the set of Caterpillar Decomposition paths.

**Spherical Embedding**