Introduction

Lorem Anjelo coglione ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

A new semantic similarity meausure

To tackle semantic similarity among terms, we relied on `WORDNET`.
We decided to take into consideration only the `hypernym-relationships` to determine words similarity. The key point was to find the `Lowest Common Ancestor`(LCA)in the WORDNET hypernym tree. It specifies the lowest parent two senses have in common.
Intuitively, the deeper the LCA, the stronger the similarity. This is true, but also the distance between the two terms is important.
If two pairs of terms meet at the same LCA, it does not mean they are identical in the similarity. If they live on two separate branches, then they are not so similar, and this should be taken into account.
This is why we introduced another metric, we called `Path Distance`(PD), which is the length of the path from $\texttt{t}_1$ to $\texttt{t}_2$ in the tree (for this one, the longer, the worse).
Finally our similarity measure is composed of two elements

- PD : Path Distance
- LCAD : Lowest Common Ancestor Depth

Given two terms, for each pair of senses we compute

$$CSIM(s_1, s_2) = \frac{1}{\sqrt{PD + 1}} \times \log(LCAD) \tag{1}$$

We added the square root to the first term of the equation, in order to smooth it out. We did the same for the second one. At the end

$$CSIM(t_1, t_2) = \max_{(s_i, s_j)} CSIM(s_i, s_j) \tag{2}$$

This turns out to work well also to measure the non-similarity, in fact suppose that two terms have a small `PD` because they are close each other, but live near the root of the tree. In this case, `LCAD` will lower the similarity.