## Maturity Score

Technologies are given a maturity score that gives an indication of how wide-spread the use of the technology is.

The maturity score is a number between 0 and 1 and the closer to 1 the more mature a technology is. Input to this score is the usage count, which looks for specific occurrences of a technology term that indicate that the technology is actually used. For example, take the sentence below.

*Fluorescence in situ hybridization (FISH) is a laboratory method used to detect and locate a DNA sequence*

This sentence indicates that *FISH* is a technology that is actually in use. On the other hand, the sentence below provides no such evidence.

*These fragments are on the order of 100 thousand base-pairs, and are the basis for most FISH probes.*

The usage count evidence is derived using 13 very simple patterns on the textual context of a technology term. In the implementation, patterns are feature bundles that are compared to the feature vectors of the technology terms, if the intersection of the feature bundle and the vector is non-empty then the pattern matches and the usage count for the term is incremented by one.

Patterns were created manually by selecting common features and determining whether the terms that have the feature were technologies that could be used. Typically, the most useful feature is the previous verb feature (prev\_V). There are 13 patterns for English, three of which are listed below.

|  |  |
| --- | --- |
| maturity-use: | prev\_V = use | uses | used | using |
| maturity-select: | prev\_V = select | selects | selected | selecting |
| maturity-store: | prev\_V = store | stores | stored | storing |

The pattern matching used for creating the usage count is a simple and fast process that is applied to all feature vectors for all terms. Results are tallied for each term and for each year, so we will end up with a table like this

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 2010 | 2011 | 2012 | 2014 | 2015 |
| Computer | 1243 | 432 | 986 | 1002 | 784 |
| FISH | 14 | 10 | 45 | 49 | 61 |

The maturity score is calculated only for those terms for which there are enough data, which was determined to mean that the term has to occur at least 25 times, and it is calculated for each year using the following algorithm:

1. Get the rough usage count for the term.
2. Adjust the count relative to the size of the data set for the year. For example, if we have only two years, 2010 and 2011, and we have twice as much data for 2010, then the usage count for 2011 would be doubled.
3. Let's call this adjusted count c. We now take log(c+1)/log(highest\_count), where highest\_count is the highest number of matches for any term in the year (with the table above we would be using the count for computer). We take log(c+1) to make sure that a count of 0 leads to a 0 score instead of an error and we use log(highest\_score) to make sure our highest value is 1.

Most maturity scores are low and most maturity scores are 0, reflecting that we found no evidence of the term being used. The graph below shows the distribution of scores over 600,000 US patents, which resulted in 353,577 terms with maturity scores.