Build a semantic space by hand

The question

Douglas Adams, *Mostly harmless*

The major difference between a thing that might go wrong and a thing that cannot possibly go wrong is that when a thing that cannot possibly go wrong goes wrong it usually turns out to be impossible to get at or repair.

- **Question:** Produce a semantic space with the following features:
  - Eliminate all function words prior to processing (determiners, prepositions, punctuation, coordinations)
  - Word window of +/- 2 words around the target.
  - Rows (targets): difference, wrong, thing, go
  - Columns (contexts): major, difference, go, wrong, thing.
  - Weighting function: $w(x,y) = \frac{freq(x,y)}{freq(x)freq(y)}$

1- Delete function words

Douglas Adams, *Mostly harmless*

major difference thing might go wrong thing cannot possibly go wrong is thing cannot possibly go wrong goes wrong usually turns be impossible get repair

2- Set up semantic space matrix

	major	difference	go	wrong	thing
difference					
wrong					
thing					
go					

3- Produce frequency counts

Douglas Adams, *Mostly harmless*

major difference thing might go wrong thing cannot possibly go wrong is thing cannot possibly go wrong goes wrong usually turns be impossible get repair

major 1	go 3	thing 3
difference 1	wrong 4	

4- Identify windows around target words

Douglas Adams, *Mostly harmless*

major difference thing might go wrong thing cannot possibly go wrong is thing cannot possibly go wrong goes wrong usually turns be impossible get repair

major **difference** thing might
major difference **thing** might go
thing might **go** wrong thing
might go **wrong** thing cannot

go wrong **thing** cannot possibly
cannot possibly **go** wrong is
possibly go **wrong** is thing
wrong is **thing** cannot possibly

cannot possibly **go** wrong goes
possibly go **wrong** goes wrong
wrong goes **wrong** usually turns

5- Fill in the co-occurrence matrix

	major	difference	go	wrong	thing
difference	1	0	0	0	1
wrong	0	0	3	2	2
thing	1	1	2	2	0
go	0	0	0	3	2

6- Apply weighting function

	major	difference	go	wrong	thing
difference	1	0	0	0	$\frac{1}{3}$
wrong	0	0	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{6}$
thing	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{6}$	0
go	0	0	0	$\frac{1}{4}$	$\frac{2}{9}$

Cosine similarity

- Cosine similarity between two vectors is defined as:

$$\cos(A, B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \quad (1)$$