**MODELO VECTORIAL: SIMILITUD COSENO**

Considérese una colección de 1.000 documentos entre los cuales se encuentran los siguientes:

Doc1: **programmers** build **computer** **software**

Doc2: most **software** has **bugs**, but good **software** has less **bugs** than bad **software**

Doc3: some **bugs** can be found only by executing the **software**, not by examining the source **code**

Los términos a considerar se han indicado en negrita.

Se pide calcular la similitud coseno entre la consulta “**computer software programmers**” y cada uno de los documentos (esquema de pesado lnc.ltc). En la tabla se indica el *df* de cada término considerado. Se han calculado los resultados truncando a dos decimales.

DEFINICIONES:









Esquema de pesado lnc.ltc:

* para los **documentos** log-pesado, no idf y normalización coseno;
* para la **consulta** log-pesado, idf y normalización coseno.