This project was built in a period of 4 weeks. The purpose is to aggregate news content from a variety of publishers, such as The Sun, BBC etc. It communicates with the google search engine with common news links phrases (e.g. “Guardian business news”). These links are then opened and using ‘HTMLAgilitypack’ to de-serialize the HTML and retrieve article links. The articles are opened, gathering information related to the stories (i.e. article contents, title, header). A word similarity algorithm identifies articles written about the same story using cosine similarity [Figure 1]. A grouping algorithm further validates the news based on low and high threshold. This algorithm is implemented under the Models->FilterLogic->Identical word similarity folder.

I picked this project as it demonstrates a variety of software engineering practices and programming styles. It is a project that uses the model-view-viewmodel architectural pattern to separate the business and back-end logic. SOLID design principles make the code maintainable and readable. Elements of objected oriented programming allow for re-usability and maintenance. Functional programming, utilizing recursion and LINQ to traverse data collections. Communication with the database is made to store articles related to the same story for retrieval for the front-end (IOS and Android).

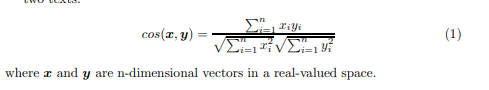


Figure 1