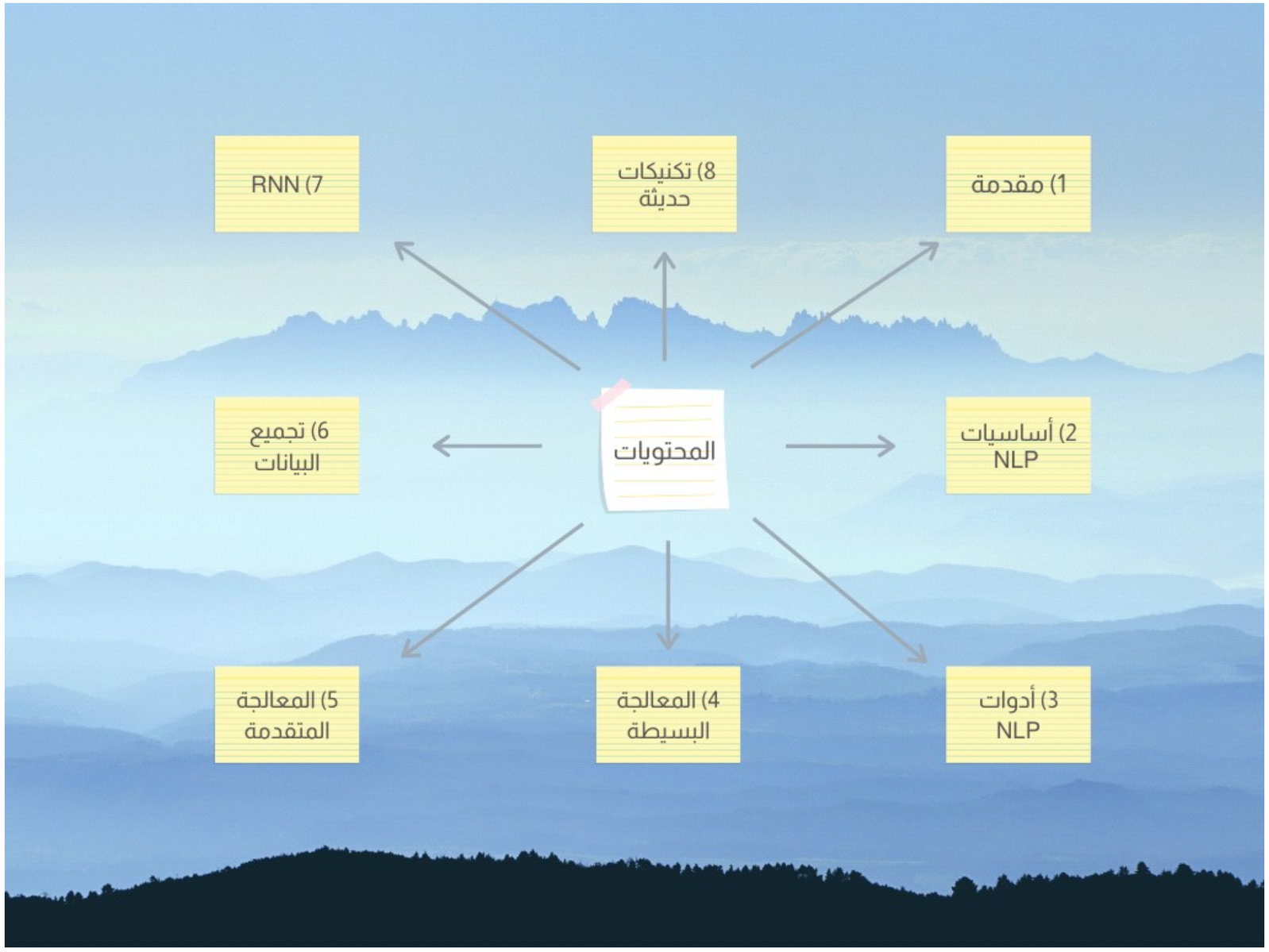
Natural Language Processing المعالجة اللغوية الطبيعية



المحتويات

1) مقدمة	المحتويات	ما هو NLP	تاريخ NLP	العقبات و التحديات	التطبيقات				
2) أساسيات NLP	المكتبات	الملفات النصية	ملفات pdf	البحث في النصوص					
3) أدوات NLP	Tokenization	Sent. Segm.	POS	Stem & Lemm	NER	Stopwords	Matchers	Syntactic Struc.	T.Visualization
4)المعالجة البسيطة	Word Meaning	Word embed	T. Vectors	Word2Vec	BOW	TF-IDF	Text Similarity	Dist. Similarity	
5)المعاجلة المتقدمة	T. Classification	T. Clustering	LDA	NMF	L. Modeling	GloVe	Lexicons	NGrams	T. Generation
	Sent. Analysis	Naïve Bayes	Vader	Auto Correct	Ans. Questions			Summarization & Snippets	
6)تجميع البيانات	Tweet Collecting	Data Scraping	Information Extraction		Information Retrieval			Relative Extraction	Search Engine
7) RNN	Seq to Seq	LSTM	GRU	Rec NN\TNN					
8)تكنيكات حديثة	Word Cloud	CNN	T. Forcing	Attention Model	Transformer	Bert	FastText	Gensim	Chat Bot

القسم الأول : مقدمة

الجزء الأول : المحتويات

==========

و هنا عدد من المفاهيم الهامة , التي يجب أن نراجع عليها سريعا :

- **تجميع البيانات النصية collecting text data**

و هي التي تتم عبر قراءة الملفات النصية , سواء كانت بامتداد txt , pdf , csv , xlsx

و اثناء القراءة يمكن ان نقوم بعمليات بحث داخلها , واستخلاص بيانات محددة و ليست كلها

- **معالجة النصوص**

و هي التي تتم عبر مكتبات عديدة مثل nltk , spacy , re و التي تقوم بجميع عمليات البحث و المعالجة المطلوبة

- **جذور الكلمة stemming & lemmatization**

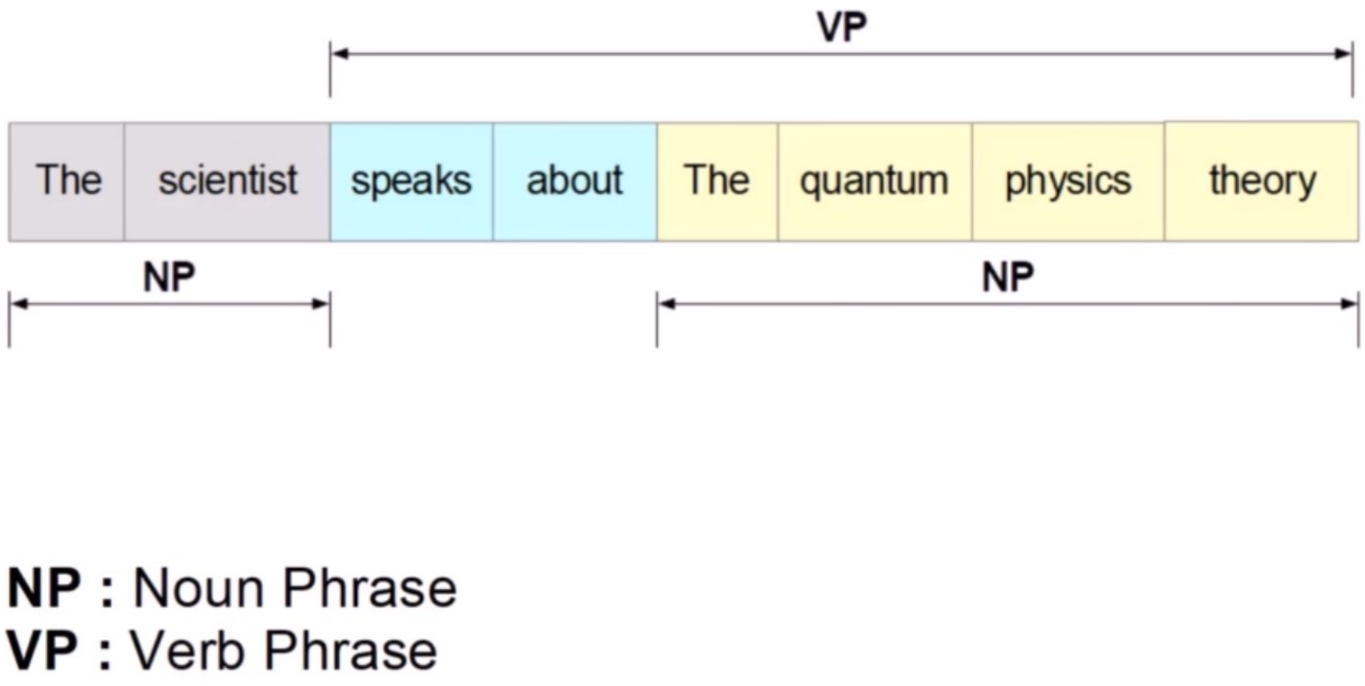
و هي خطوات ايجاد اصل الكلمة و جذرها , و حذف كل الاضافات عليها ,و ايجاد الفعل الاصلي لها

- **جزء النصوص POS**

و هي تحديد نوع هذه الكلمة في قواعد اللغة , هل هي اسم او فعل او صفة , و ربطها بباقي الجمل

- **تقطيع الجمل chunking**

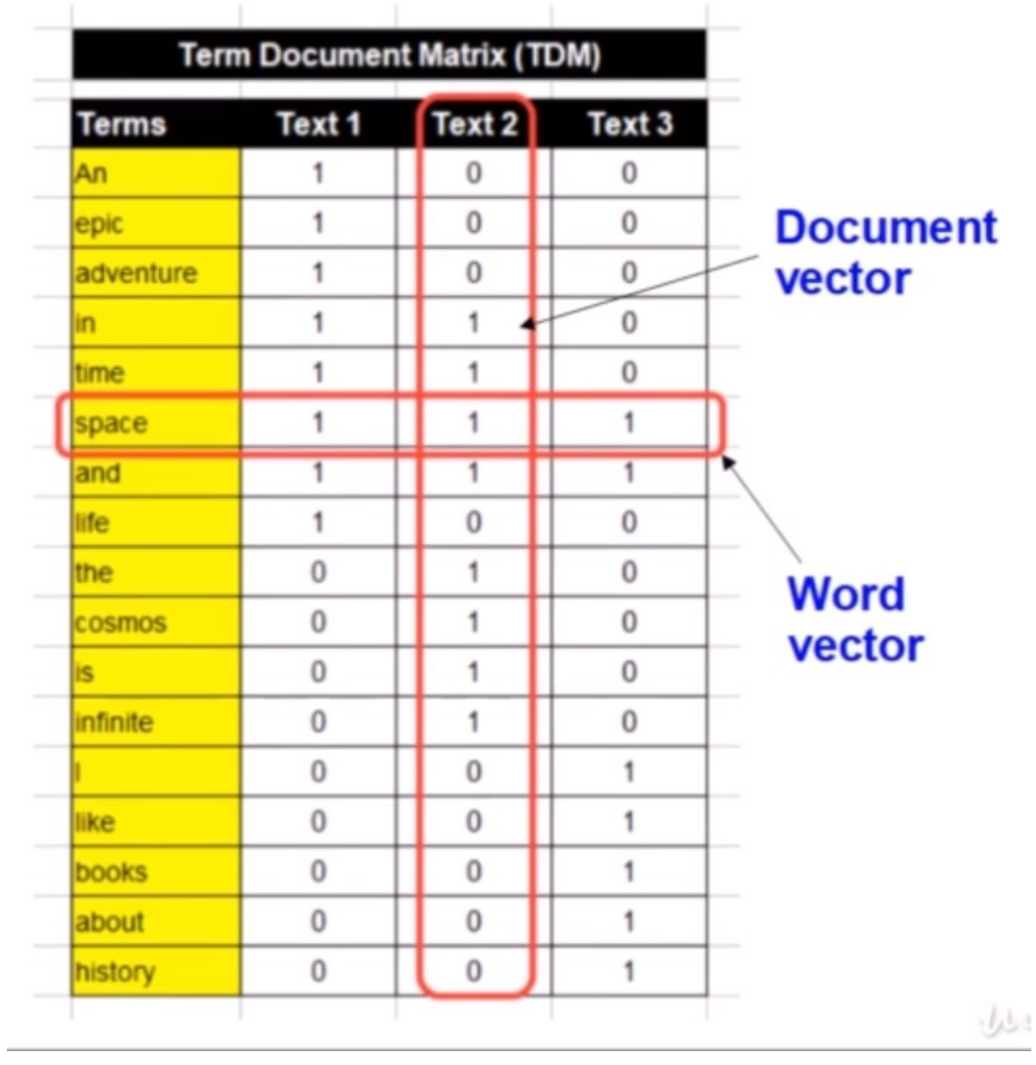
و هي تحديد اي كلمات تنتمي لبعضها البعض , و كيف تقوم بتكوين الجمل معا



NP : Noun Phrase
VP : Verb Phrase

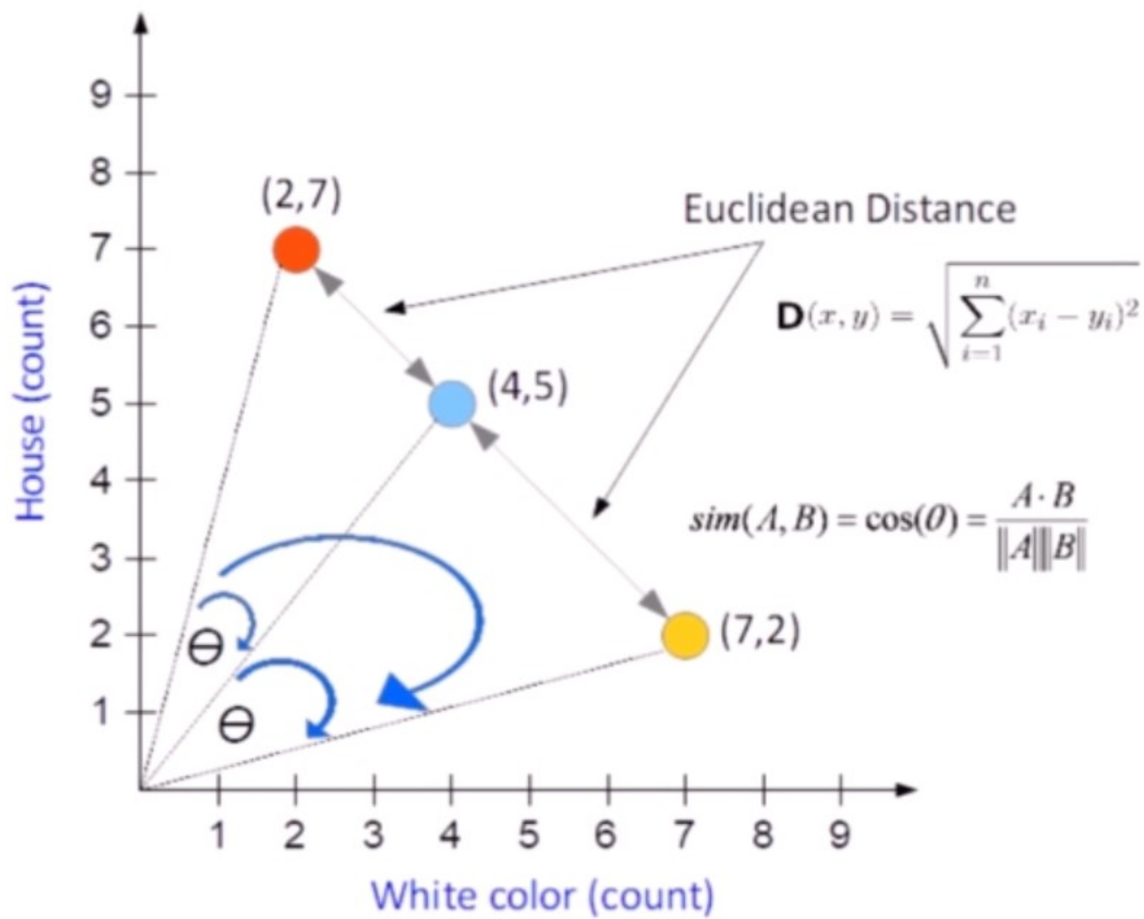
- **جعبة الكلمات Bag of Words**

و هي طريقة جمع كل الكلمات الموجودة في النص , ثم تحديد هل الجملة الاولي او القطعة الاولي فيها هذه الكلمات ام لا



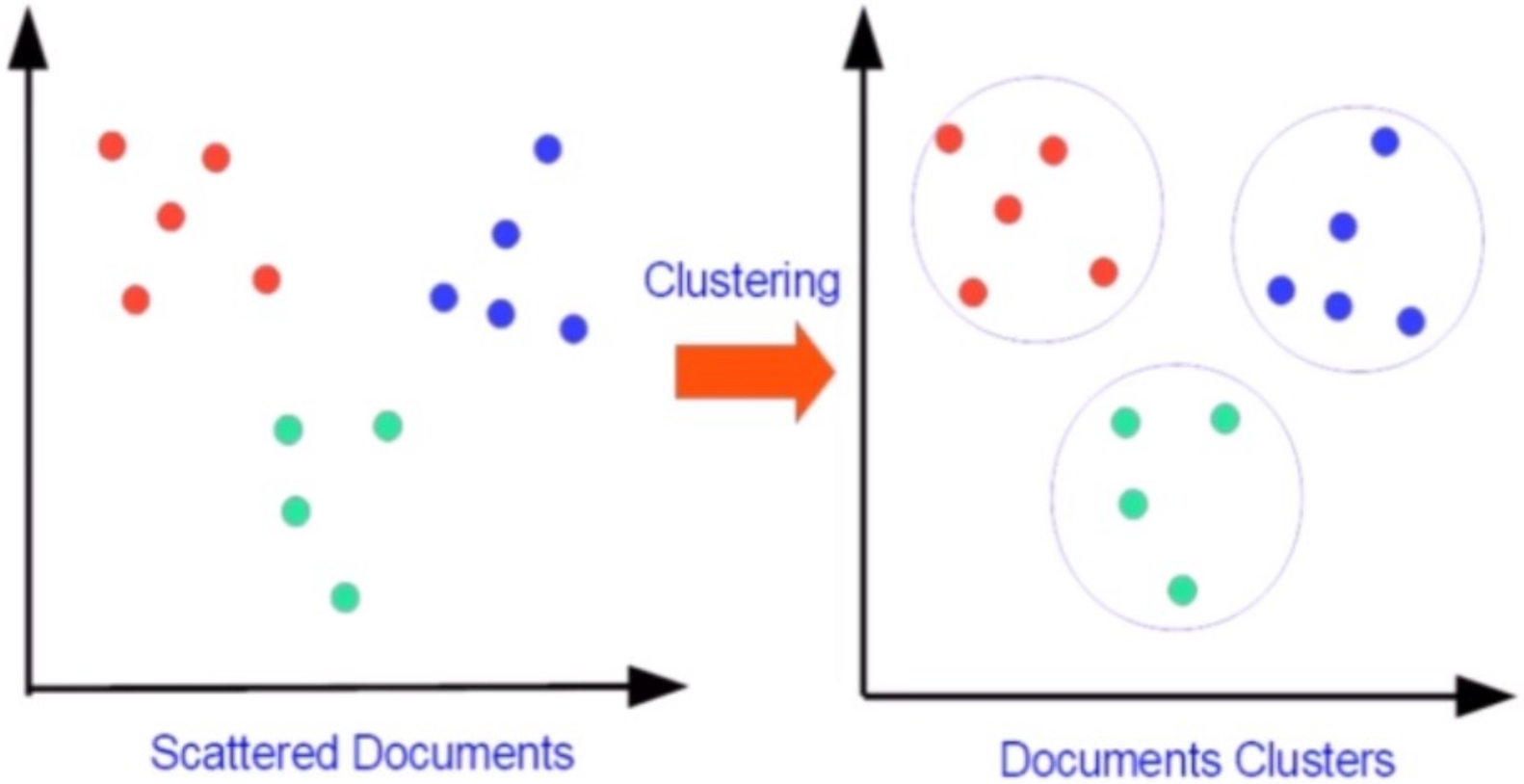
- **تشابه الكلمات text similarity**

و هي تشير الي مدي اقتراب معاني الكلمة الاولي من الثانية , عبر المعاني المضمنة embedding , ويتم حسابها باكثر من طريقة منها cos angle , او المسافة المطلقة بفيثاغورث



- **عناقيد الكلمات :**

وهي استخدام تعليم بدون اشراف لتصنيف الكلمات الي عناقيد , بناء علي معانيها المضمنة

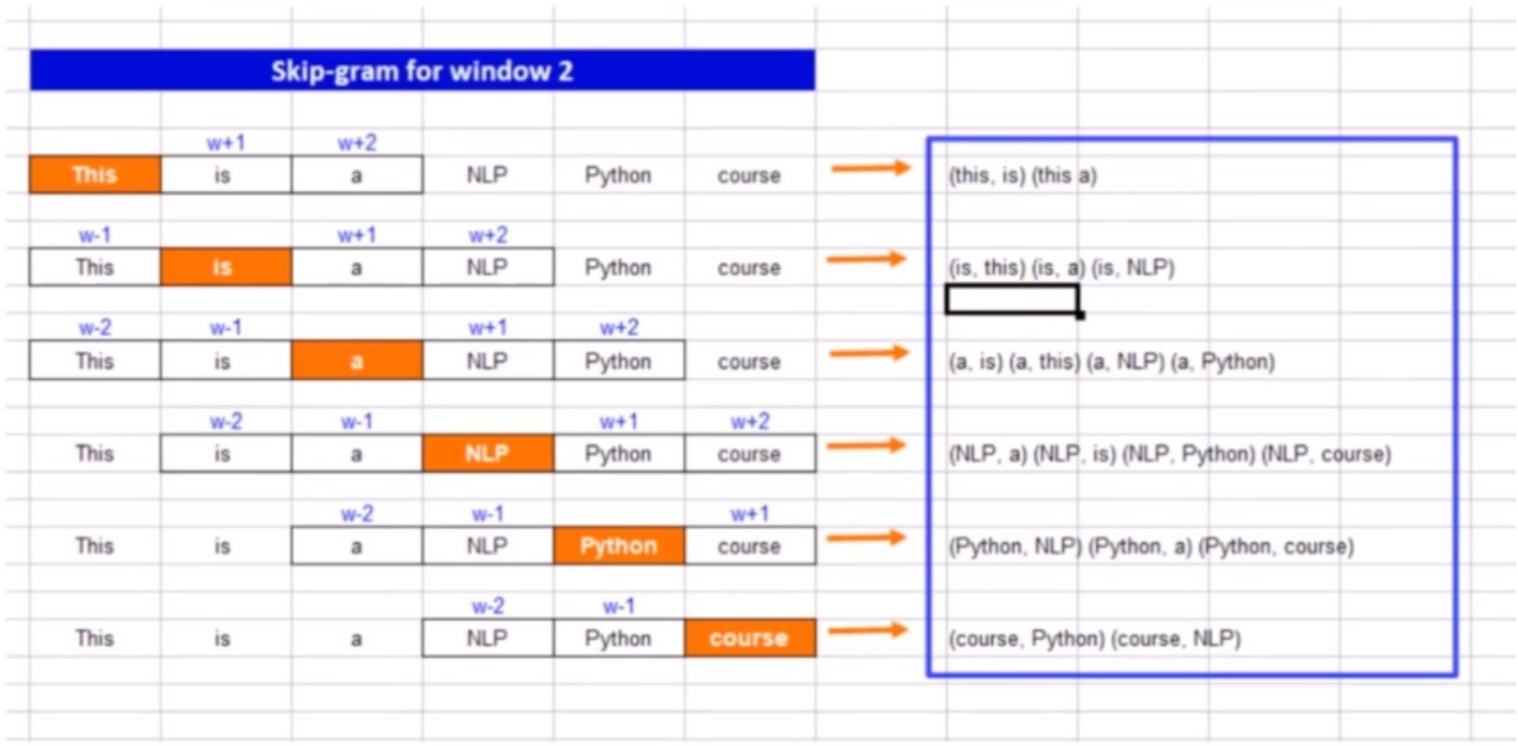


- **تكنيك LDA :**

و هي احد اساليب التعليم بدون اشراف , حيث نقوم بتقسيم الجمل الي عدد من الاقسام , بناء علي المعاني الدفينة في الجمل

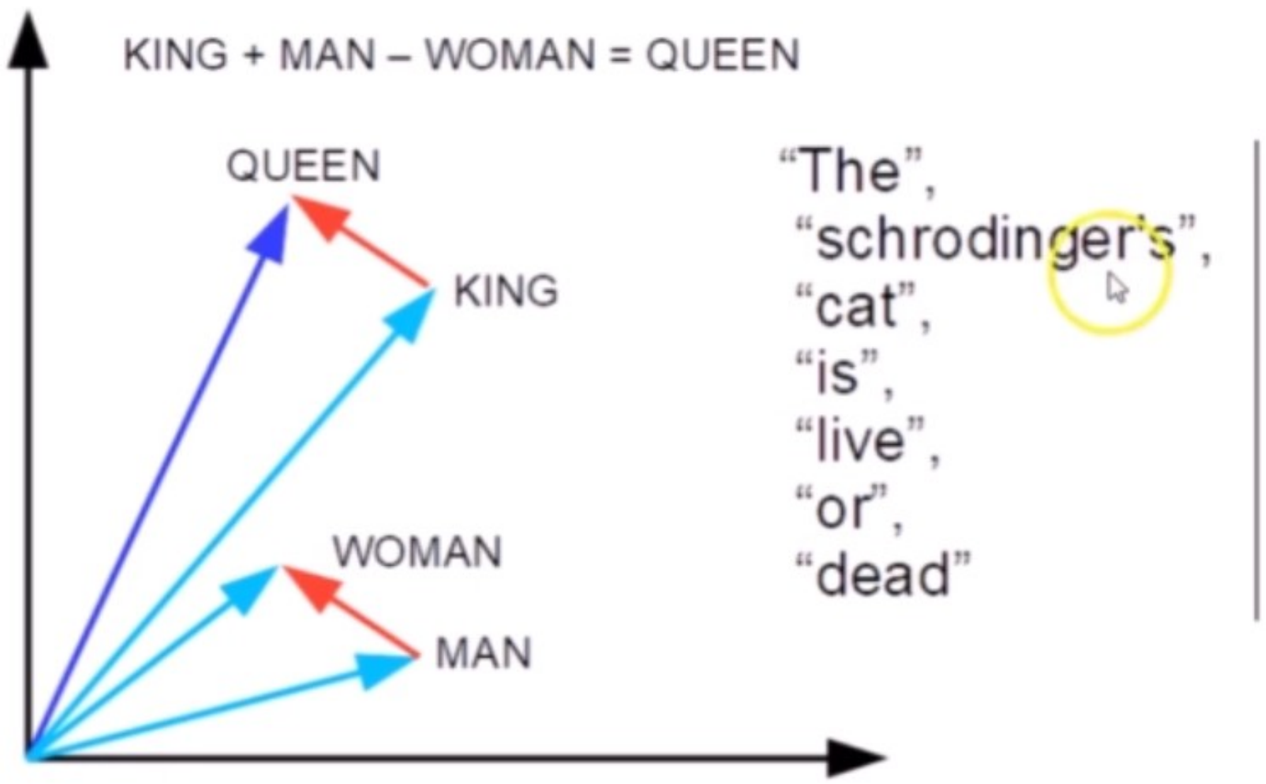
- **تكنيك NGrams :**

- و هو اسلوب التعامل مع عدد معين من الكلمات السابقة و الكلمات التالية , لاستنتاج الكلمة الجديدة , او للتصنيف



- **تضمين الكلمات :**

وهو اظهار ال vectors الخاصة بكل كلمة , والتي يمكن من خلالها معرفة مدي اقتراب او ابتعاد الكلمات



Word embeddings are vectors that represent multiple features of