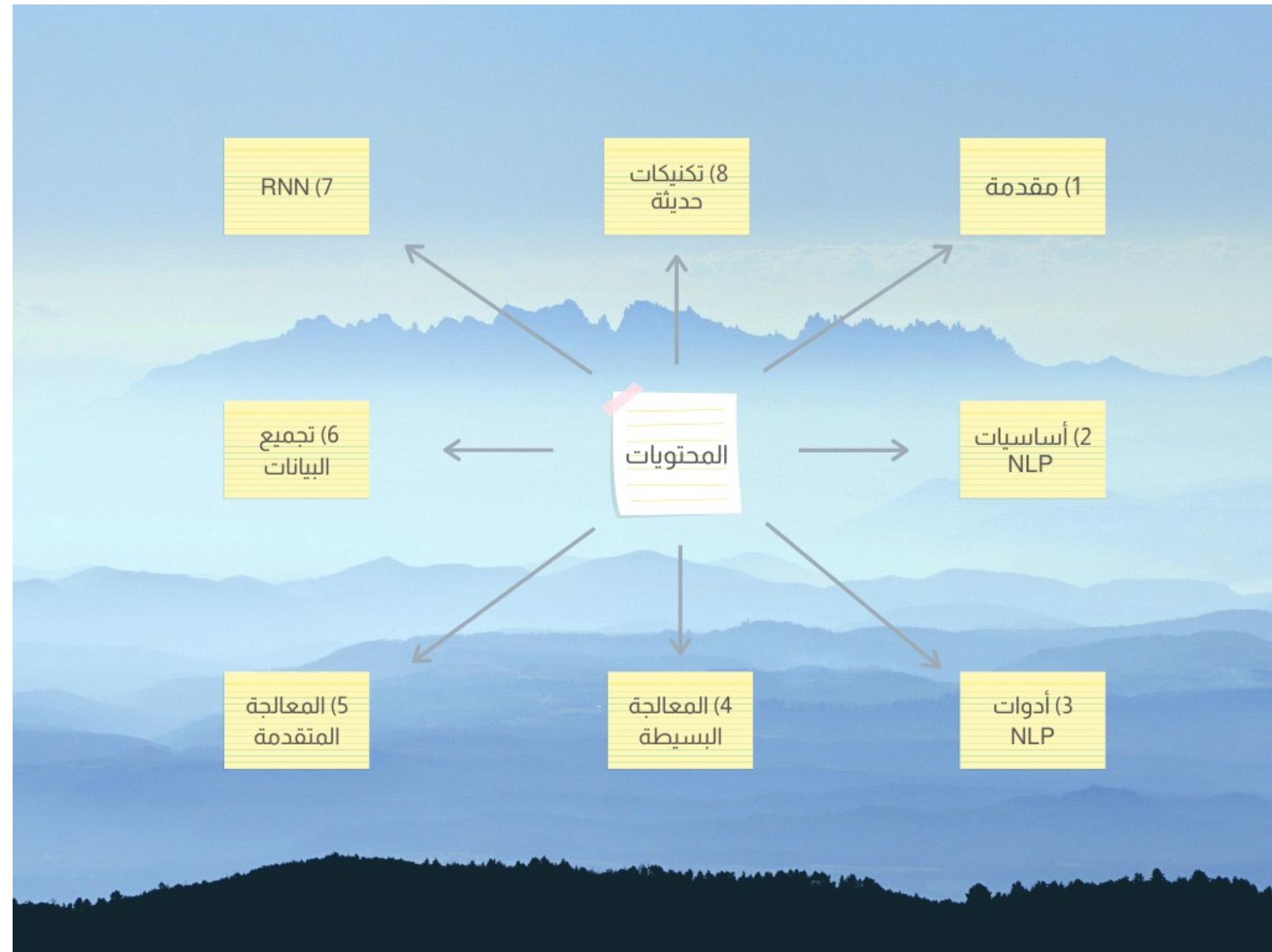


NATURAL LANGUAGE PROCESSING

المعالجة اللغوية الطبيعية



المحتويات

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

القسم الأول : مقدمة

الجزء الأول : المحتويات

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

- **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 **re , spacy , nltk** و التي تقوم بجميع عمليات البحث و المعالجة المطلوبة

• جذور الكلمة stemming & lemmatization

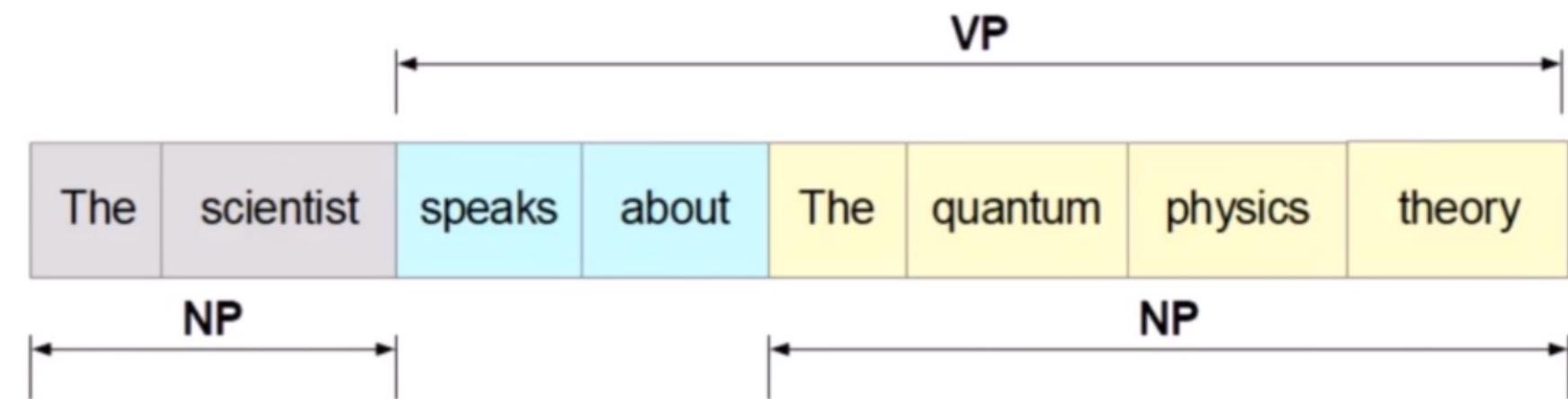
و هي خطوات ايجاد اصل الكلمة و جذرها , و حذف كل الاضافات عليها , و ايجاد الفعل الاصلي لها

• 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

• جعبه الكلمات Bag of Words

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

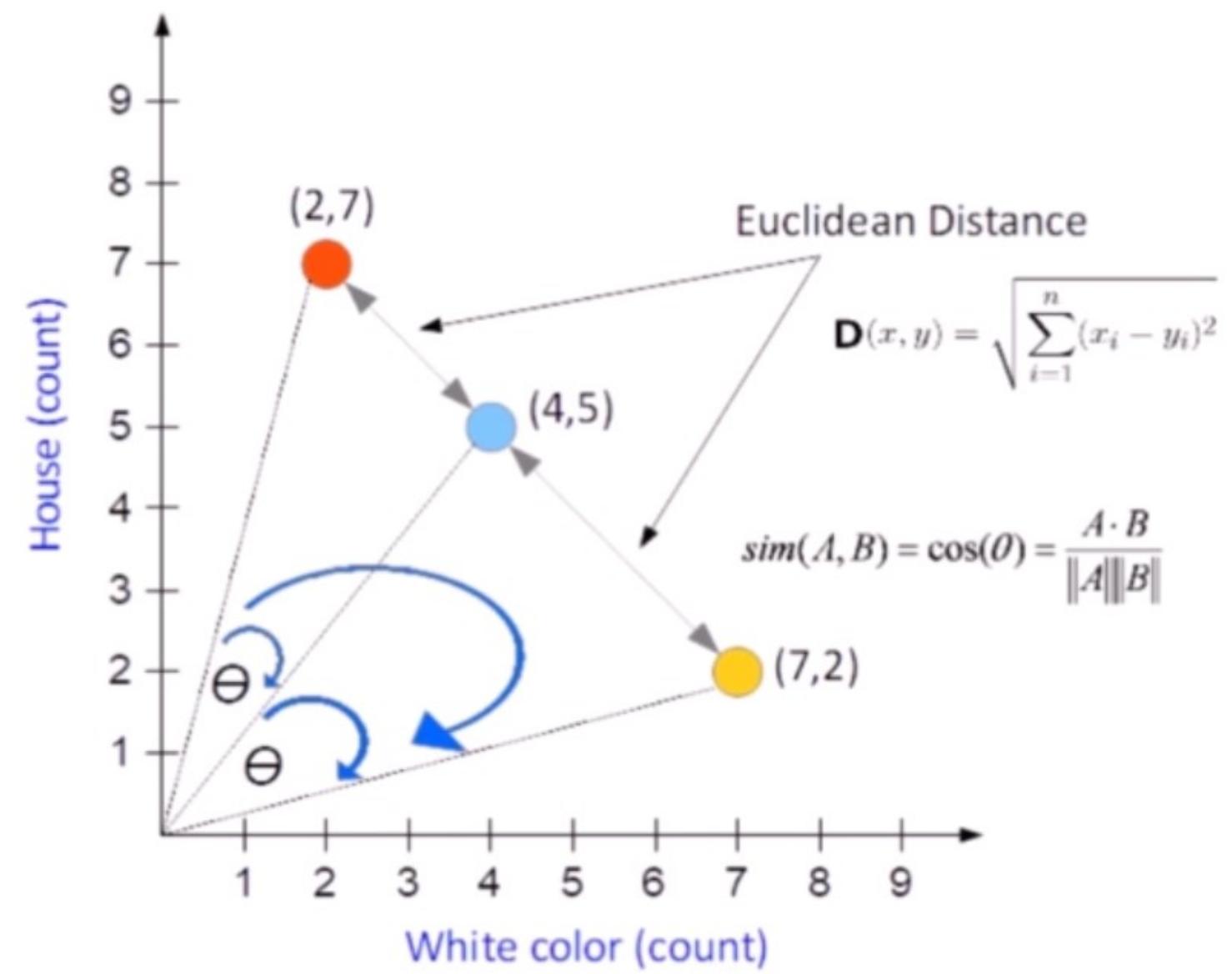
Term Document Matrix (TDM)			
Terms	Text 1	Text 2	Text 3
An	1	0	0
epic	1	0	0
adventure	1	0	0
in	1	1	0
time	1	1	0
space	1	1	1
and	1	1	1
life	1	0	0
the	0	1	0
cosmos	0	1	0
is	0	1	0
infinite	0	1	0
I	0	0	1
like	0	0	1
books	0	0	1
about	0	0	1
history	0	0	1

Document vector

Word vector

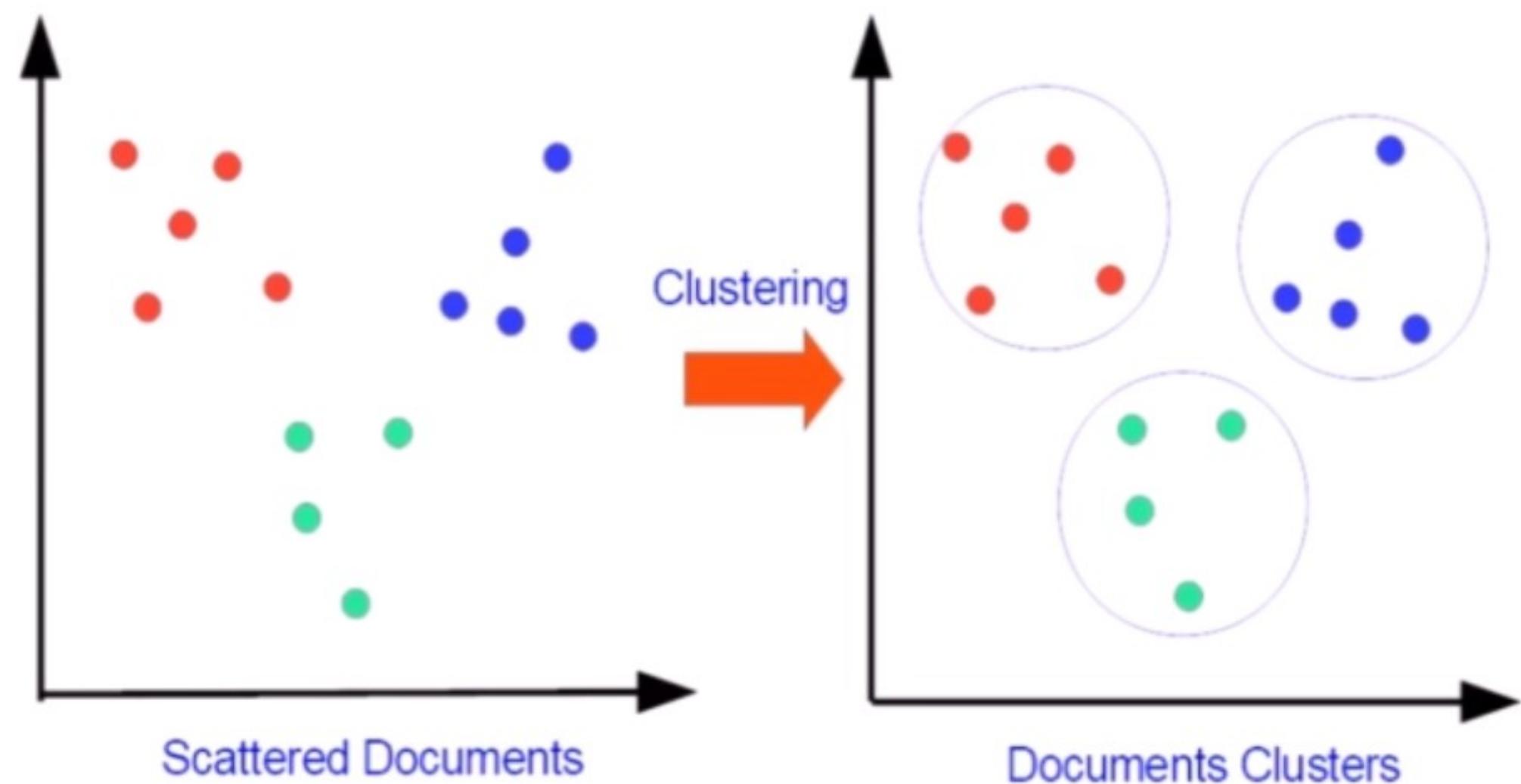
• تشابه الكلمات text similarity

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



- عناقيد الكلمات :

وهي استخدام تعليم بدون اشراف لتصنيف الكلمات الى عناقيد ، بناء على معانيها المضمنة

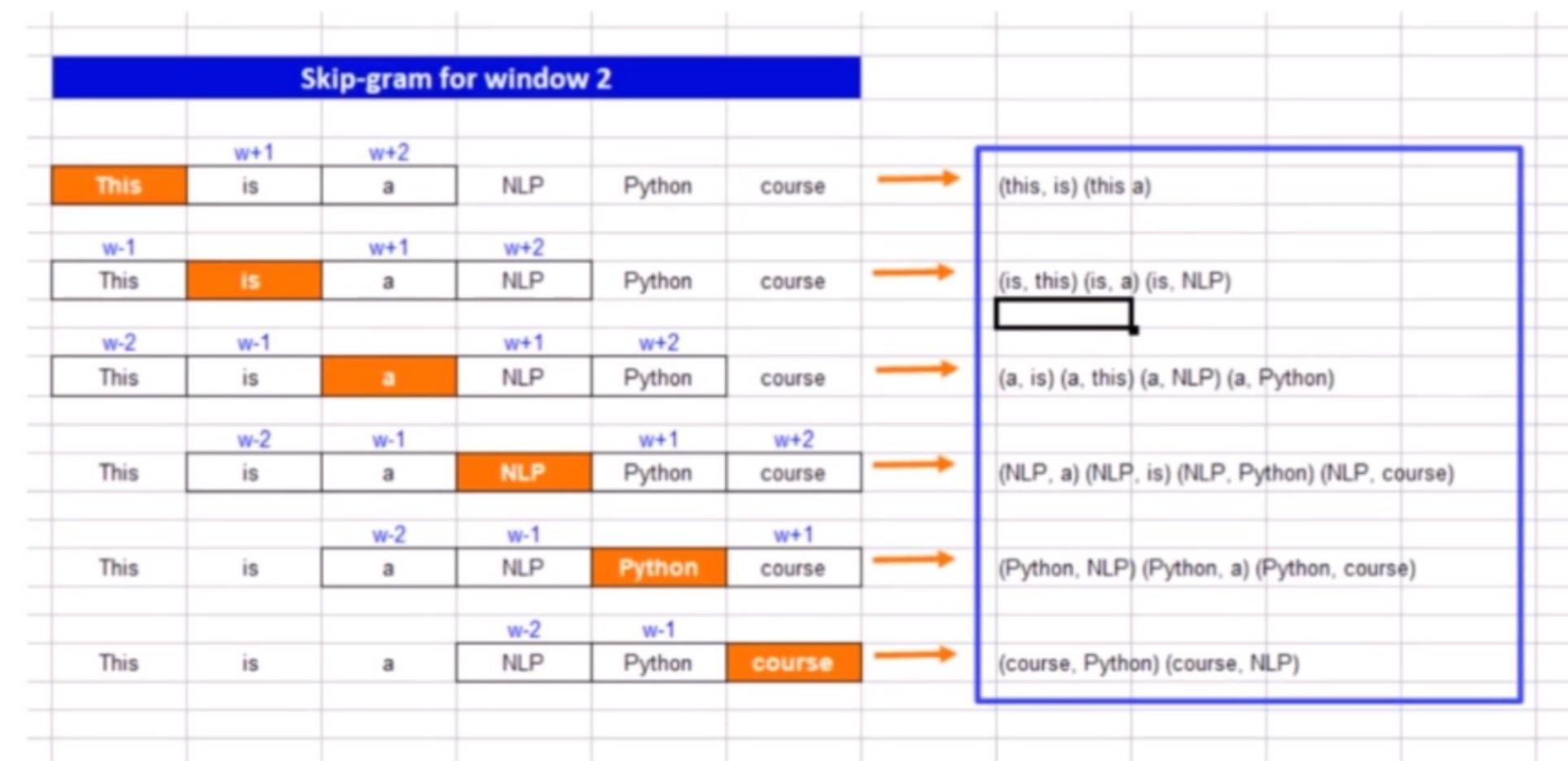


• تكنيك : LDA

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

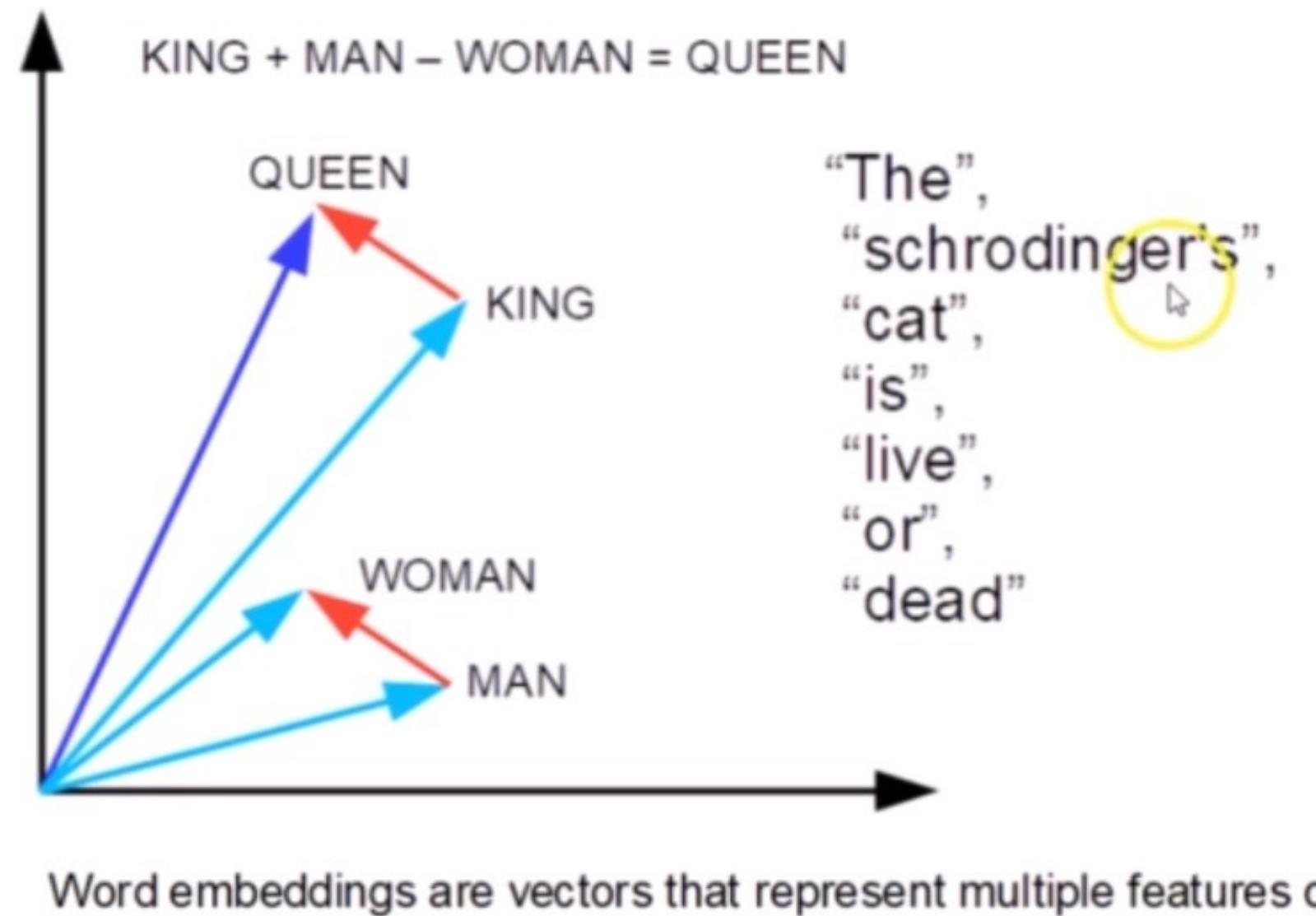
• تكنيك : NGrams

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



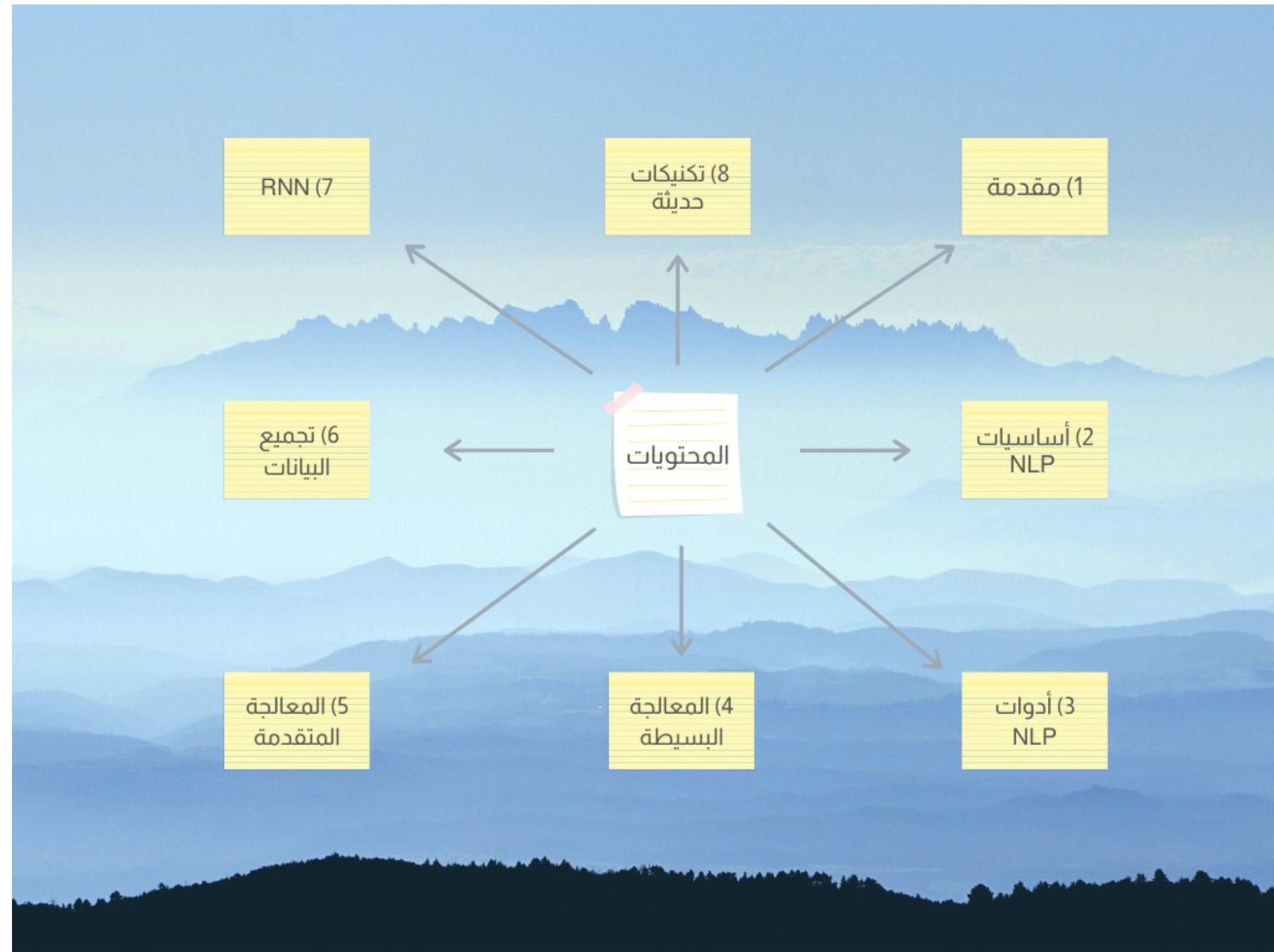
- **تضمين الكلمات :**

وهو اظهار ال vectors الخاصة بكل كلمة ، والتي يمكن من خلالها معرفة مدى اقتراب او ابعاد الكلمات



NATURAL LANGUAGE PROCESSING

المعالجة اللغوية الطبيعية



المحتويات

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

القسم الأول : مقدمة

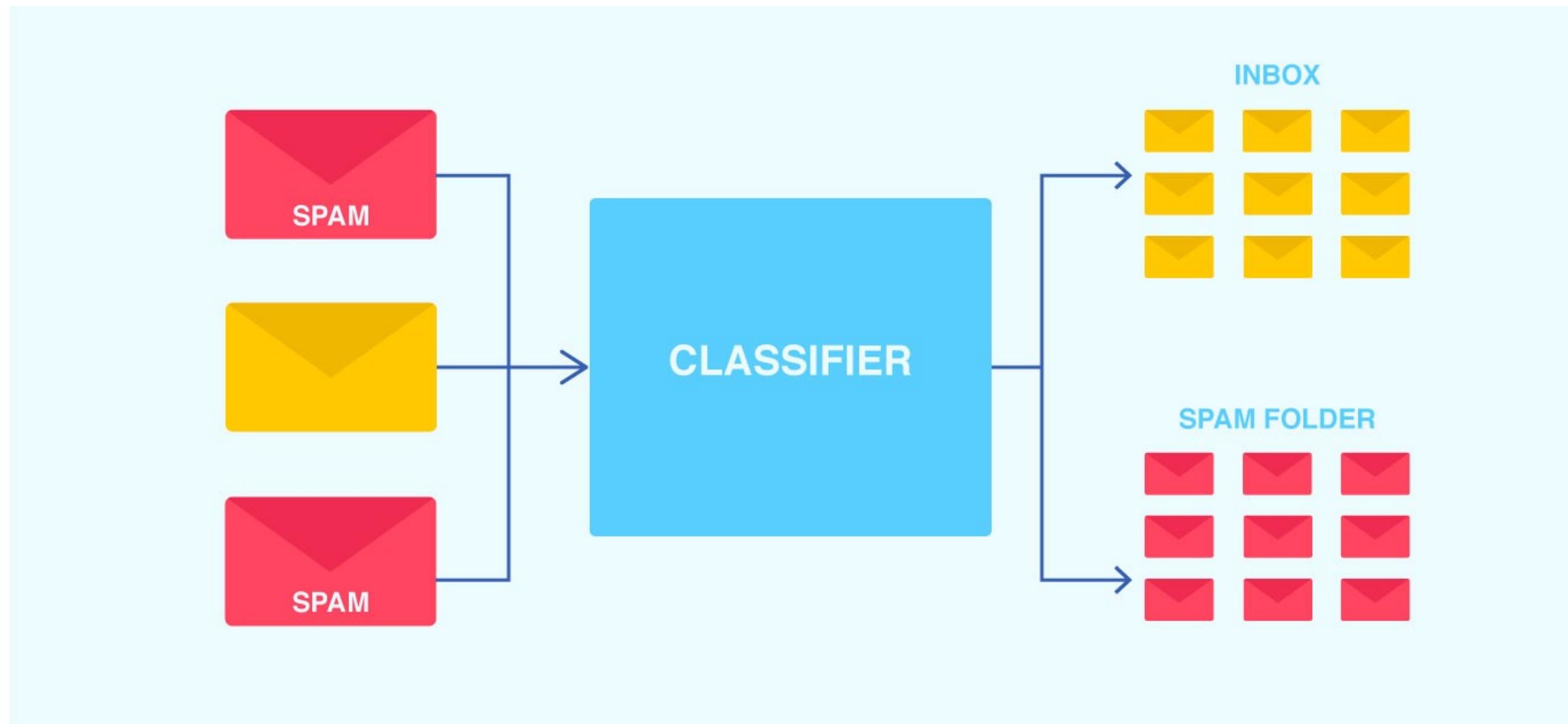
الجزء الثاني : ما هو NLP

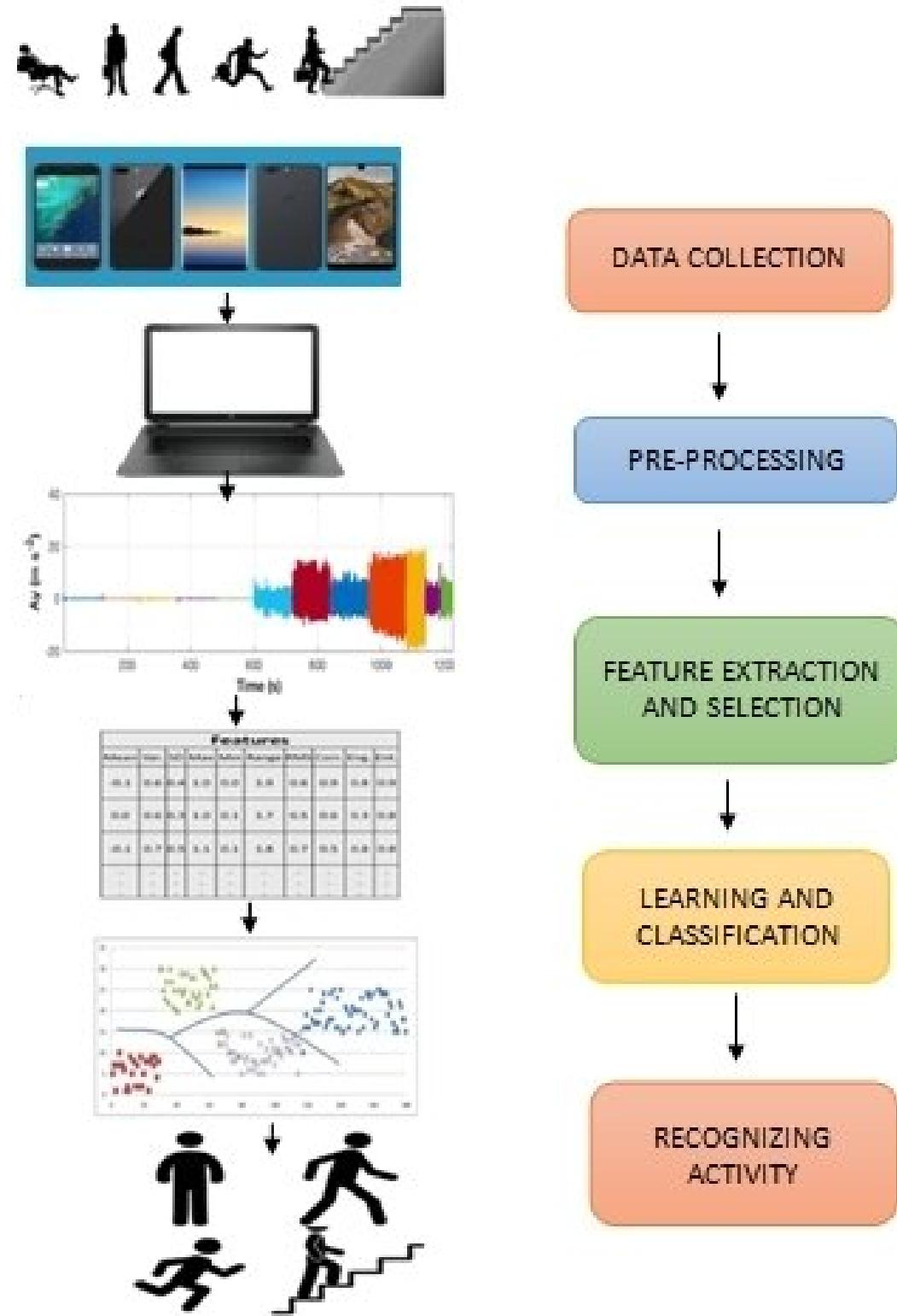
ما هو المعالجة اللغوية الطبيعية : NLP

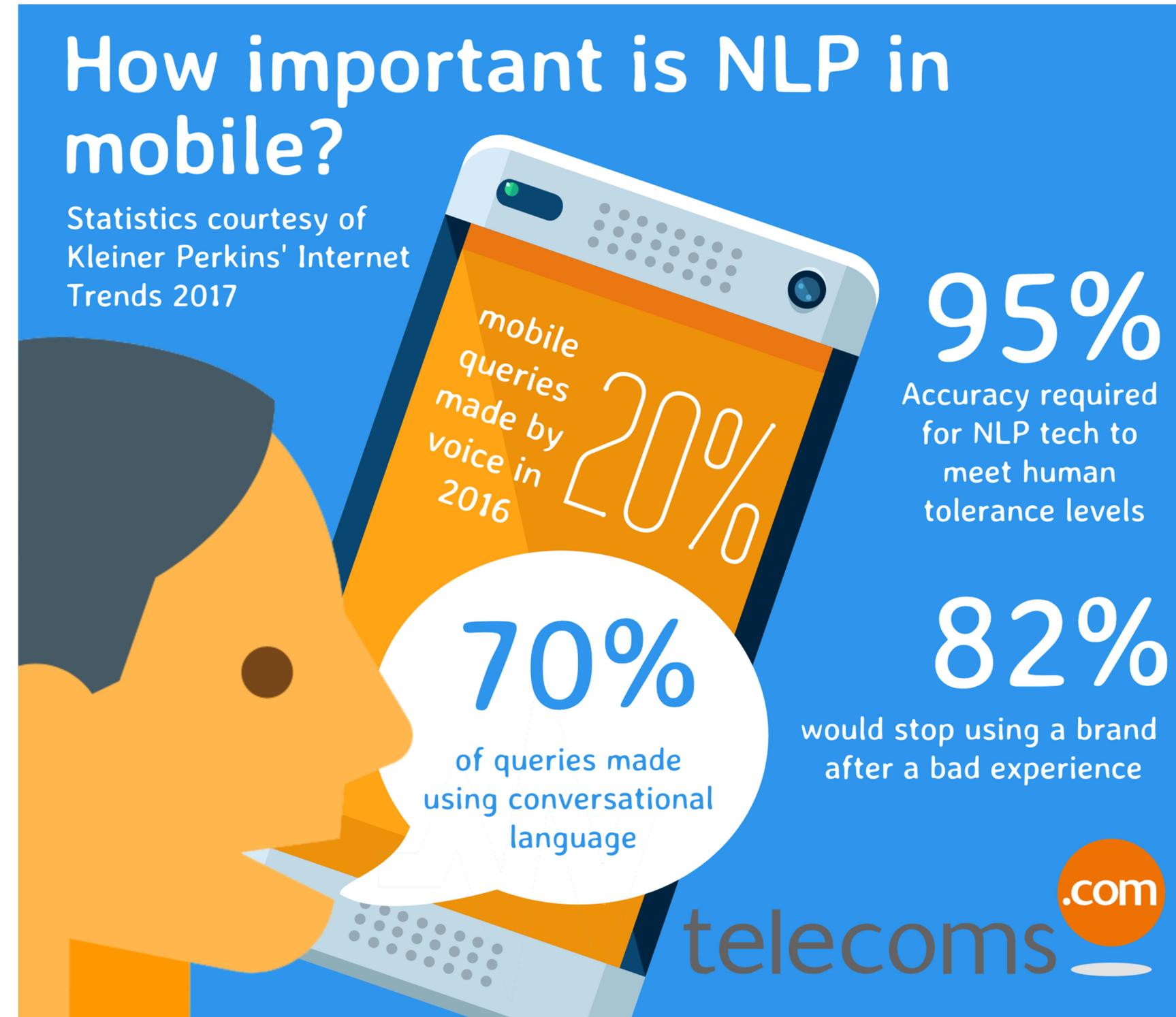
is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data.

فرع من علوم اللغويات و الذكاء الاصطناعي ، وهو المختص بالتفاعل و التواصل بين اجهزة الكمبيوتر ، والإنسان الطبيعي ، وكذلك لتناول و تحليل التعامل الغوي و النصوص ذات الكميات الكبيرة .

و هي ليست لها علاقة بعلم NLP وهو Neuro-Linguistic programming







تسارع استخدام الانترنت على مستوى العالم

