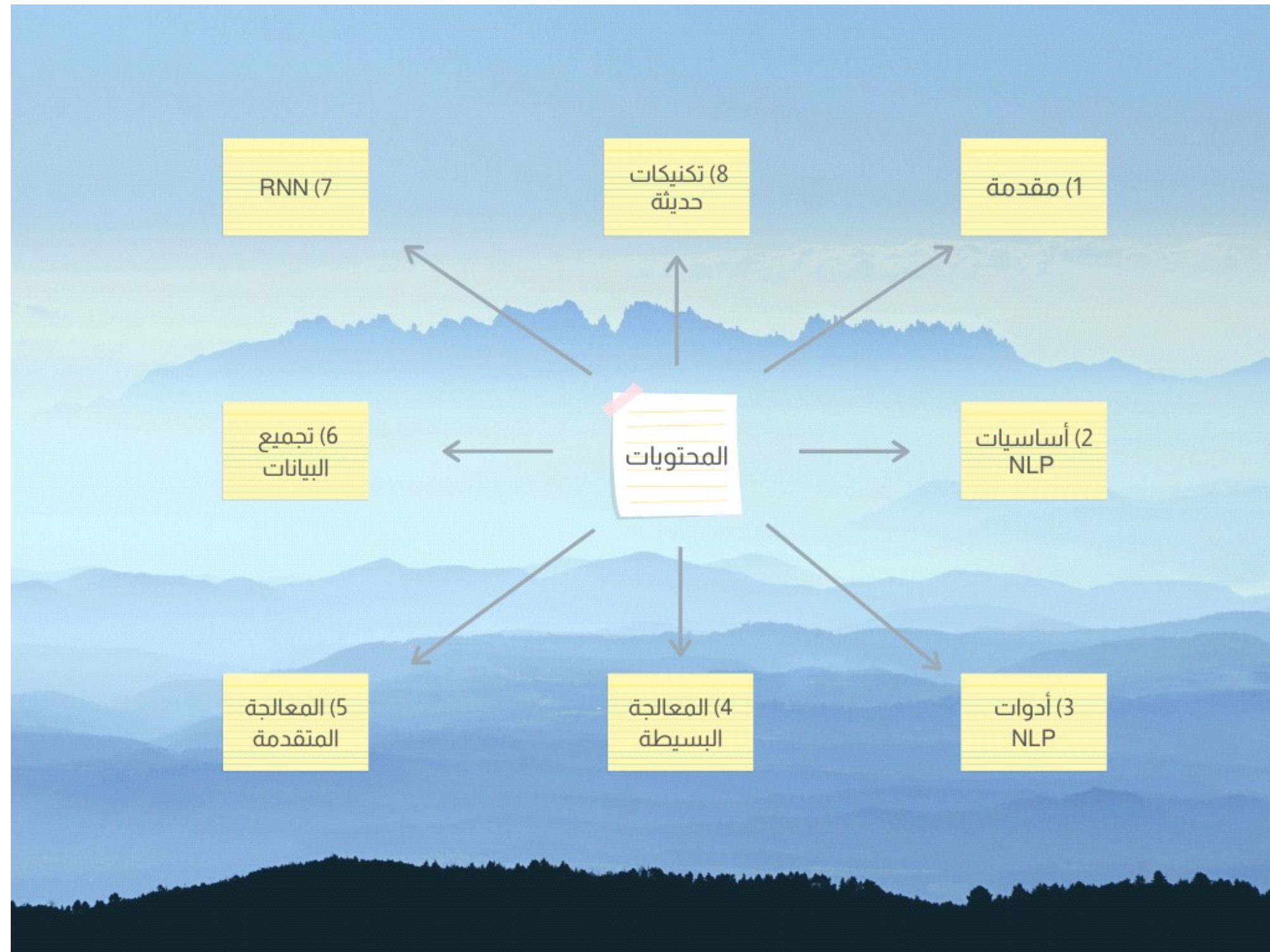


NATURAL LANGUAGE PROCESSING

المعالجة اللغوية الطبيعية



المحتويات

				التطبيقات	العقبات و التحديات	تاريخ NLP	ما هو NLP	المحتويات	1) مقدمة
					البحث في النصوص	ملفات pdf	الملفات النصية	المكتبات	2) أساسيات NLP
T.Visualization	Syntactic Struc.	Matchers	Stopwords	NER	Stem & Lemm	POS	Sent. Segm.	Tokenization	3) أدوات NLP
	Dist. Similarity	Text Similarity	TF-IDF	BOW	Word2Vec	T. Vectors	Word embed	Word Meaning	4) المعالجة البسيطة
T. Generation	NGrams	Lexicons	GloVe	L. Modeling	NMF	LDA	T. Clustering	T. Classification	5) المعالجة المتقدمة
	Summarization & Snippets		Ans. Questions		Auto Correct	Vader	Naïve Bayes	Sent. Analysis	
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) تكتيكات حديثة

القسم الرابع : المعالجة البسيطة للنصوص

الجزء الأول : Word Meaning

=====

نتناول هنا مصطلحات هامة في معاني الكلمات , ومرادفاتها للتعرف علي المصطلحات المختلفة في هذا المجال :

- هيئة Lemma Form :

○ هي الكلمة بعد ان يتم ارجاعها إلي جذورها : sing , run , book

*_**

- هيئة wordform :

○ هي الكلمة بعد ان تم اضافة الاضافات عليها : sung , runner , bookings

*_**

- نوع المعني :
 - وهي المعاني المختلفة التي يمكن ان تكون لنفس الكلمة :
- ...a **bank**¹ can hold the investments in a custodial account...
 - "...as agriculture burgeons on the east **bank**² the river will shrink even more"

* * * * *

● كلمات Homographs :

- الكلمة التي يكون لها نفس الحروف , لكن أكثر من معني : bank , like , book , serve
- ليس بالضرورة ان تكون المعاني بعيدة عن بعض , لكن اغلب الكلمات لها معاني مختلفة قليلا مثل : run , drive
- هناك ما يسمى اختبار زيوجما zeugma لمعرفة هل استخدام الكلمة هنا بنفس المعني ام لا
- فكرة الاختبار تقوم علي ايجاد جملتين بالمعنيين المختلفين :

Ahmed runs 2 km daily

Ahmed runs the restaurant

- ثم ضم الجملتين معا :

Ahmed run 2 km daily the restaurant

- و نري هل الجملة متسقة ؟ , اذا لم تكن متسقة فالمعاني مختلفة
- مثال مختلف :

Afaf wrote a book

Afaf wrote the message

Afaf wrote the book & the message

* * * * *

● کلمات : Homophones

- الكلمات التي يكون لها نفس النطق , لكن بحروف مختلفة , ومعاني مختلفة : write:right , piece:peace
○ وهذه الكلمات تتسبب أحيانا في مشكلة مع خوارزم voice to text , حيث يكون النطق متطابق , و يتم التفريق بينها في المعني و السياق فقط

* * * * *

● کلمات : Synonyms

- و هي الكلمات التي تحل محل بعضها البعض في أغلب المواضع , مثل :

Big/large , car/automobile , water/H₂O

- و يتم الإشارة الي مدي تقارب الكلمتين , بناء علي كمية المرات التي يمكن استبدال الكلمتين مكان بعضهما البعض في كل الجمل

- فکلمتي car/vehicle ذات ارتباط ضعیف
- و کلمتي big/large ذات ارتباط متوسط
- و کلمتي convince/persuade ذات ارتباط قوی

○ و لكن يصعب جدا العثور علي كلمتين متطابقتين , و ذلك بسبب المعاني الاخري التي تستخدم لها الكلمات , فكلكتي big/large لها الكثير من التشابه في الحجم , لكن big تستخدم ايضا مع big brother اي الاخ الاكبر , بينما large لا يمكن استخدامها في ذات الموضع

* * * * *

● مفهوم : Antonyms

○ و هي الأضداد , أي الكلمات التي لها عكس معنى كلمة أخرى , في منظور معين

Dark/light , short/long ○

○ هناك كلمات لها أكثر من معكوس حسب استخدامها

○ فكملة big عكسها small اذا تكلمنا عن الحجم

○ و نفس الكلمة عكسها little اذا تكلمنا عن السن

* * * * *

● مفهوم Hyponym & Hypernym

- وهو مفهوم الأصول Hypernym و الفروع Hyponym
- فالسيارة تعتبر Hyponym من المركبات
- الفراولة تعتبر Hyponym من الفواكه
- مفهوم Hypernym هو العكس اي الاصل
- أي أن المركبات هي Hypernym للسيارة
- الفواكه هي Hypernym للفراولة

* * * * *

● و تستخدم هذه المفاهيم في :

- بناء المعاجم lexicons و المكتبات سابقة التدريب GloVe
- في Sentimental Analysis
- التعرف علي قيمة تقريبية لمصفوفة التضمين embedding matrix لمعني كلمة مجهول
- صياغة رد علي سؤال معين في chatbot
- محركات البحث , لتحديد كلمات البحث عبر استخدام كلمات معينة

* * * * *