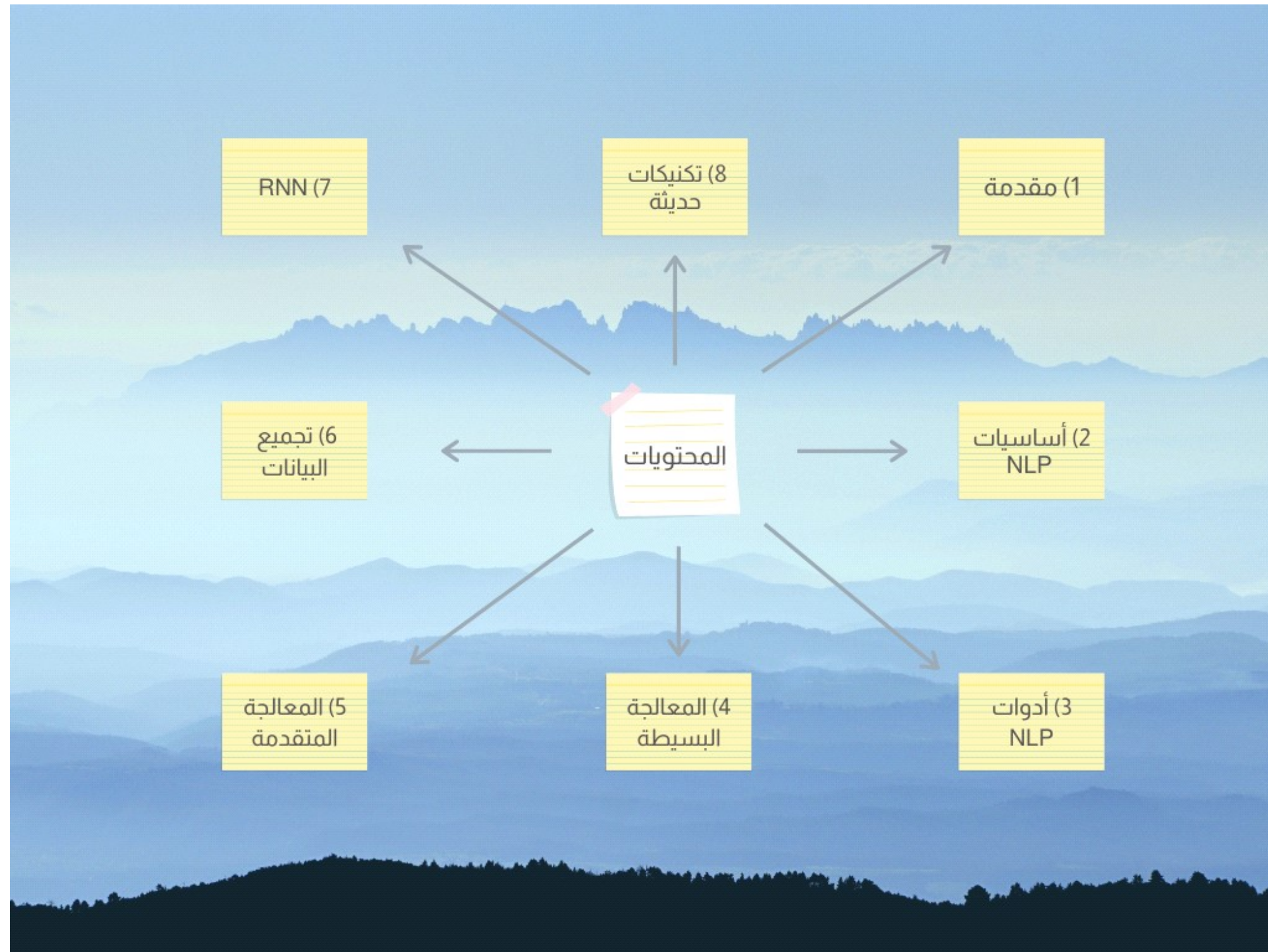


# NATURAL LANGUAGE PROCESSING

# 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	L. Modeling	NGrams	Lexicons	GloVe	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 السادس : تجميع البيانات

### الجزء الرابع : Information Retrieval

Information Retrieval  
نتناول الآن أحد طرق الـ data scraping و هي خاصة بالبحث عن النصوص في الأماكن المتاحة

يقصد بها هي البحث عن المواد المطلوبة ( غالبا ما تكون المستندات ) للبيانات غير المهيكلة ( غالبا النصوص ) , لتقوم بتلبية الاحتياجات المطلوبة , من عدد كبير من الملفات ( علي الكمبيوتر او اونلاين )

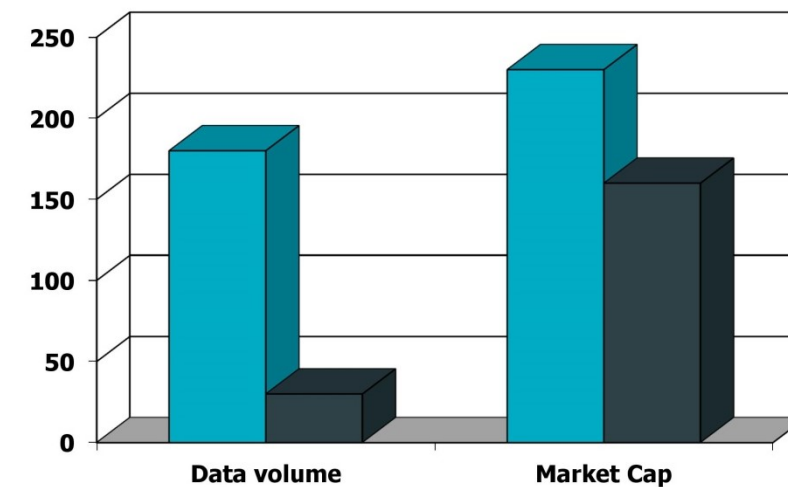
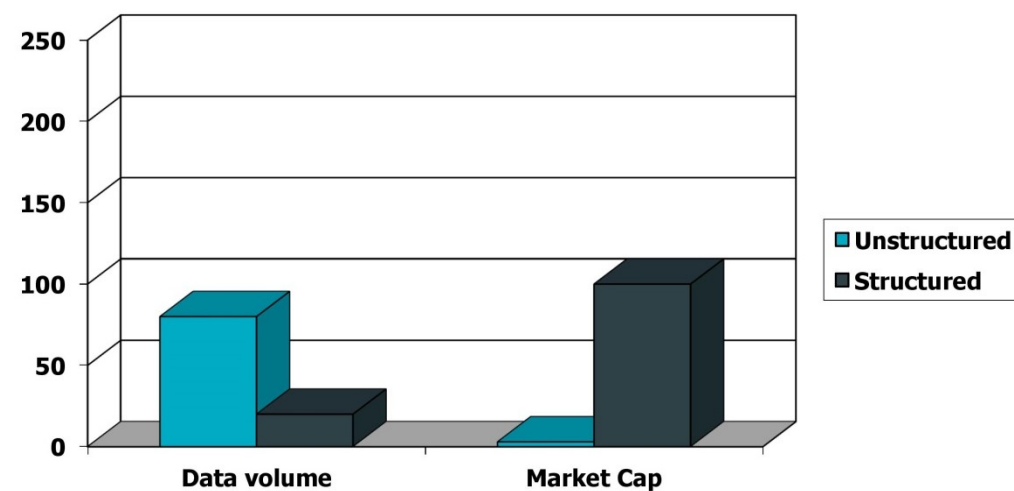
Information Retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).



و يكون البحث عبر المواقع , او الايميلات , او البحث في الحاسب , في مواقع الشركات , او اماكن الارشيف المتاحة

- These days we frequently think first of **web search**, but there are many other cases:
  - E-mail search
  - Searching your laptop
  - Corporate knowledge bases
  - Legal information retrieval

و هنا توضيح لحجم البيانات الموجود , واستيعاب السوق للبيانات المهيكلة و غير المهيكلة في التسعينات و حاليا



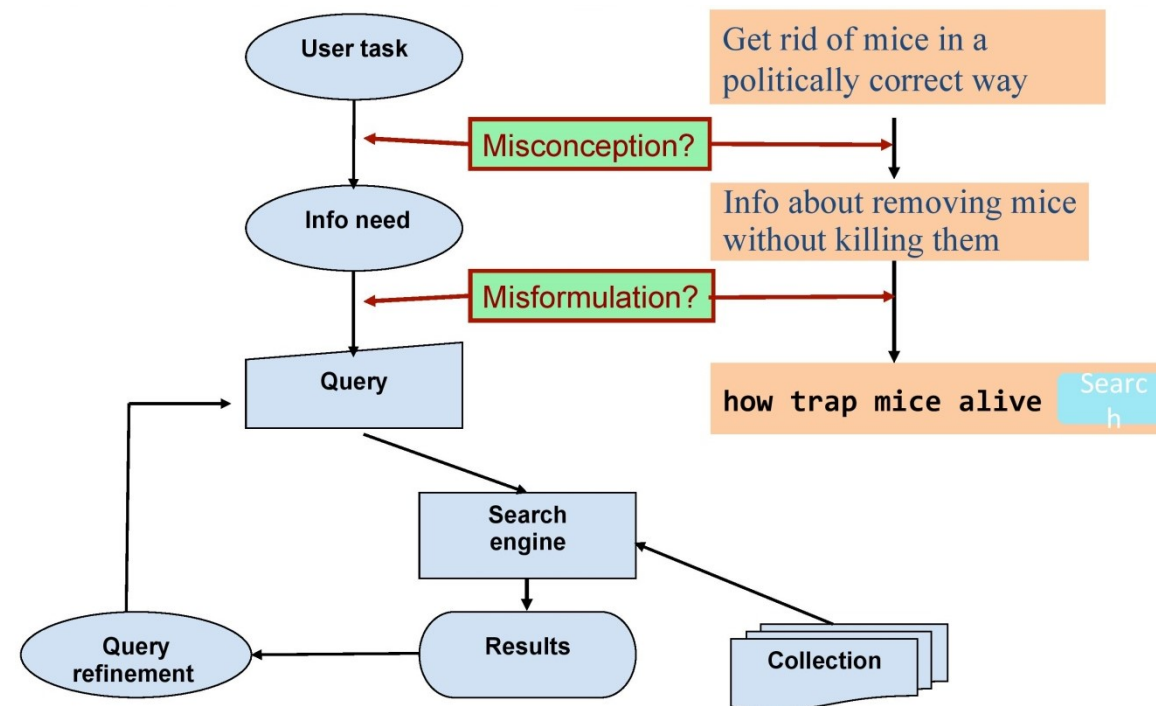
و تكون خطوات البحث كالتالي :

- أولاً تحديد المهمة : كيف نتخلص من الفئران بطريقة مناسبة
- تحديد المعلومة المطلوبة : الطريقة للتخلص من الفئران دون قتلهم
- كلمات البحث : اصطياد الفئران حية
- عمل خطوات البحث : إما أونلاين او بالبدائل السابق ذكرها
- تقييم النتائج :

○ إذا كانت كافية تم

○ إذا لم تكن : تغيير آلية البحث او كلماته

بشكل ما , هذا ما نقوم به بشكل دائم بطريقة غير شعورية



و يتم تقييم البحث كالتالي :

● الدقة : بالنسبة الي المعلومات التي حصلنا عليها , كم منها مفيد لنا

● الاستدعاء : ما هي نسبة المعلومات الهامة التي حصلنا عليها من المعلومات الموجودة في المصدر

# How good are the retrieved docs?

- *Precision* : Fraction of retrieved docs that are relevant to the user's **information need**
- *Recall* : Fraction of relevant docs in collection that are retrieved
  - More precise definitions and measurements to follow later

\* \* \* \* \*