

# NATURAL LANGUAGE PROCESSING

# المعالجة اللغوية الطبيعية



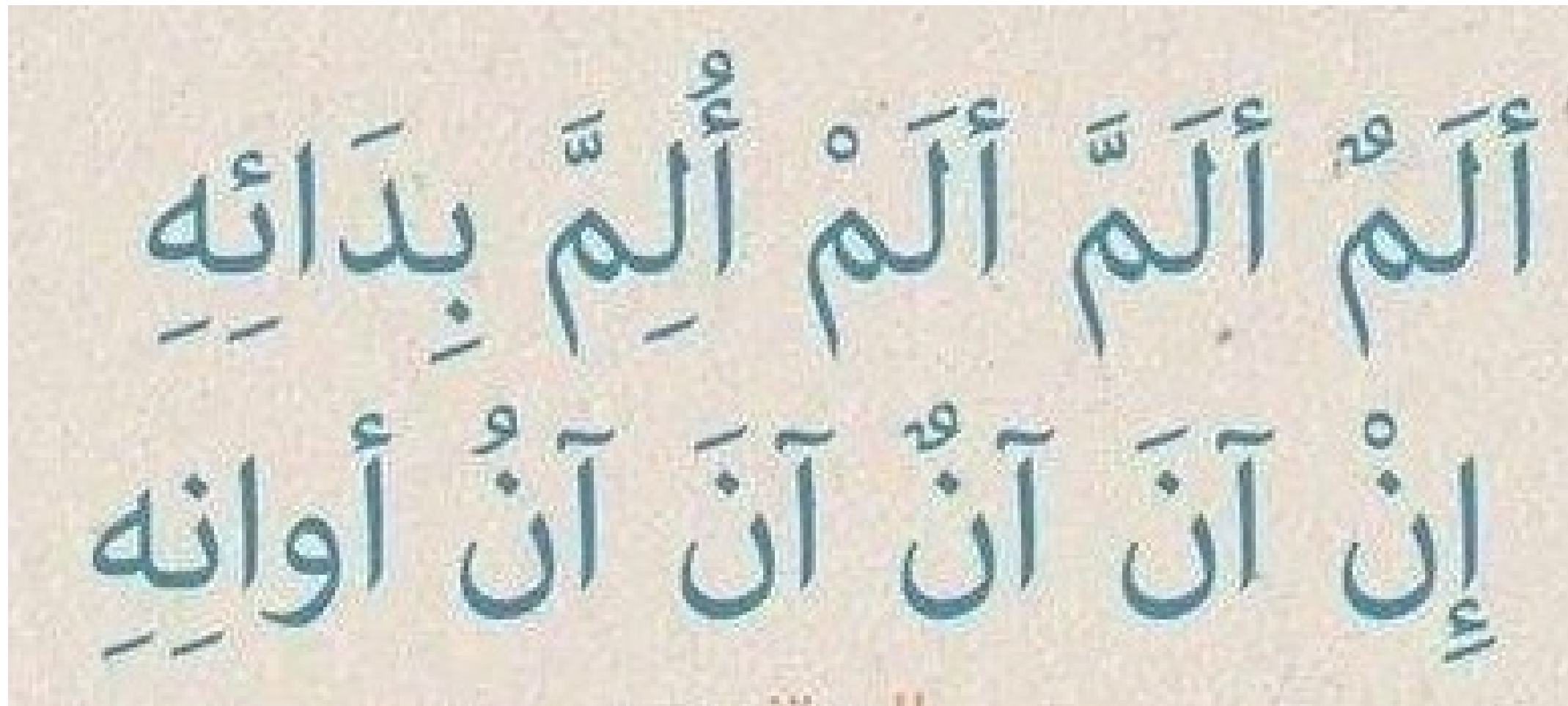
# المحتويات

				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) تكنيكات حديثة

## القسم الأول : مقدمة

## الجزء الرابع : العقبات و التحديات

## قال المتنبي :



و هي ما توضح صعوبة فهم اللغات بشكل عام

و بالإنجليزية

All the money he had had had had no effect on the him.

## و الجملة المعقدة :

This exceeding trifling witling, considering ranting criticizing concerning adopting fitting wording being exhibiting transcending learning, was displaying, notwithstanding ridiculing, surpassing boasting swelling reasoning, respecting correcting erring writing, and touching detecting deceiving arguing during debating.

و كل هذه الجمل , تشير الى صعوبة التعامل مع اللغة بشكل عام . .

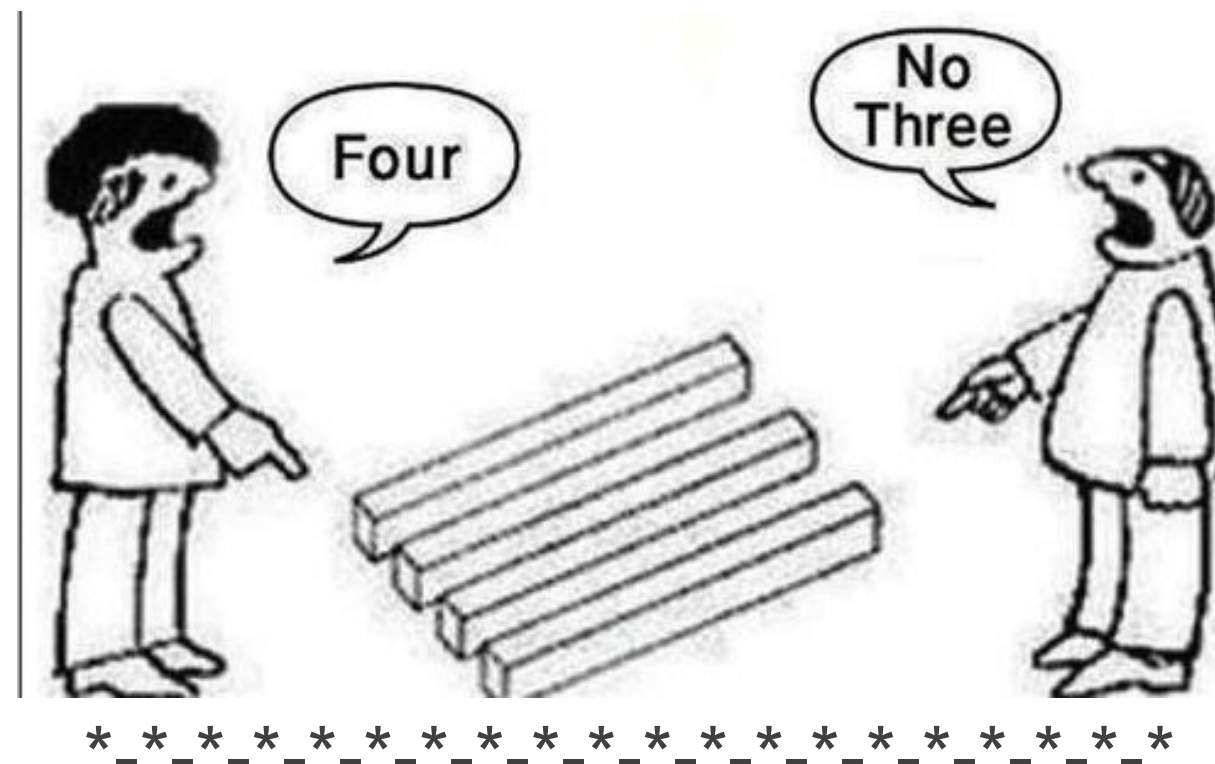
\* \* \* \* \*



يعد علم NLP من أهم و أعقد علوم ال DL حيث يتعامل مع النصوص و هي التي يساء فهمها في كثير من الأحيان بين البشر أنفسهم , لأنها تتعارض مع مبدأ Ceiling Analysis

و معضلة اجهزة الكمبيوتر انها يسهل لها التعامل مع الارقام ببساطة , والقيام بكل العمليات الحسابية عليها , لكن لا يسهل عليها التعامل مع النصوص , خاصة انها بيانات غير مهيكلة **unstructured data** و تأتي في اشكال و نماذج و لغات مختلفة

و لكن من أصعب التحديات في NLP هو ما يسمى "الزهور المحطمة crashed blossoms"



وهي تشير الى العبارات التي تكتب بطريقة معينة , فيساء فهمها من القارئ , مثال لها جملة : (مصرع السفاح , عبد  
الناصر في باكستان) حيث تم اساءة فهم الجملة عن المعني المقصود



كذلك جملة

“A woman, without her man, is nothing.”

“A woman: without her, man is nothing”.

و التي يمكن فهمها بعدة طرق

مثال آخر فجملة بسيطة مثل :

Federal raises interest rates 0.5%

يمكن ان تفهم بأكثر من طريقة هكذا :

(Federal) (raises) (interest rates) (0.5%)

(Federal raises) (interest) (rates 0.5%)

( Federal raises interest ) (rates) (0.5%)

وحتي في القرآن الكريم هناك علامات الوقوف لمنع الفهم غير السليم :

ففي سورة الأنعام (إِنَّمَا يَسْتَجِيبُ الَّذِينَ يَسْمَعُونَ وَالْمَوْتَىٰ يَبْعَثُهُمُ اللَّهُ ثُمَّ إِلَيْهِ يُرْجَعُونَ)

و في سورة المائدة يمكن ان يتم فهم الآية بمعنيين (قَالَ فَإِنَّهَا مُحَرَّمَةٌ عَلَيْهِمْ ۖ أَرْبَعِينَ سَنَةً يَتِيهُونَ فِي الْأَرْضِ ۚ فَلَا تَأْسَ عَلَى الْقَوْمِ الْفَاسِقِينَ)

\* \* \* \* \*



و في اللغة الإنجليزية هناك امثلة عديدة :





## Girl found alive in France murders car

A girl has been found hiding inside a British-registered car in which three bodies were discovered eight hours earlier, French police say.

▶ Shock at 'brutality'

Teacher Strikes Idle Kids

Red Tape Holds Up New Bridges

Hospitals Are Sued by 7 Foot Doctors

Juvenile Court to Try Shooting Defendant

Local High School Dropouts Cut in Half

و أصل التسمية يعود الي سنين بعيدة , حيث قام كاتب بنعي ابنته التي ماتت في حادث الطيران الياباني JAL بقوله :

# Violinist linked to JAL crash blossoms

أن عازفة الفيولين في الطيران الياباني قد حطمت الزهور , اي ماتت

و حينما نُشر العنوان فهم الكلام بمعنى آخر , لذا تم اتخاذ مصطلح crash blossoms يدل علي التباس الفهم

\* \* \* \* \*

## ايضا من التحديات التي تواجه الأخرى الـ NLP

- الاختلاف الكبير في اللهجات حتي في نفس اللغة : "شيخة" في مصر تختلف عن الخليج تختلف عن المغرب
- اللغة غير الرسمية المستخدمة في كثير من المحادثات
- التقطيعات غير الصحيحة في جمل
- التعبيرات المجازية : ريح شوية , اقعد علي جنب , طير من هنا
- المصطلحات الجديدة : فرمته , رنله
- الاستخدامات الساخرة للكلمات : شرفتي يا ست هانم
- كلمات متعددة الاستخدامات و المعاني : يا سلام .
- الحقائق التاريخية : محمد علي حكم مصر
- صعوبة تحديد المصطلحات الخاصة : فيلم الارهاب و الكباب جميل



# Why else is natural language understanding difficult?

## non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either♥

## segmentation issues

the New York-New Haven Railroad  
the New York-New Haven Railroad

## idioms

dark horse  
get cold feet  
lose face  
throw in the towel

## neologisms

unfriend  
Retweet  
bromance

## world knowledge

Mary and Sue are sisters.  
Mary and Sue are mothers.

## tricky entity names

Where is *A Bug's Life* playing ...  
*Let It Be* was recorded ...  
... a mutation on the *for* gene ...

But that's what makes it fun!