

# المدونات اللغوية

- التعريف والأنواع والتصميم
- اللسانيات ومعالجة اللغة الطبيعية
- المناهج المدونة اللسانية الإحصائية
- مناهج معالجة اللغة الطبيعية
- معالجة اللغة الطبيعية
- الختام: اتجاهات المشاريع

[https://www.youtube.com/watch?v=XbfjBpj2\\_g&t=927s](https://www.youtube.com/watch?v=XbfjBpj2_g&t=927s)

د. سلطان ناصر المجيلول

جامعة الملك سعود

# التعريف والأنواع والتصميم

• **التعريف:** مجموعة ضخمة من اللغة (النصوص) تُجمع وتُحفظ في قواعد بيانات أو ملفات محوسبة.

• **الأنواع:** النوع، والغرض، والعدد، والتصميم

• **التصميم:**

1. التصميم لإطار العينة:

أ. الأوعية: الصحافة، الكتب، المناهج الدراسية، الحوار، المجالات...

ب. المجال: سياسة، اقتصاد، ثقافة، طب، علوم اجتماعية...

ج. الموضوع: موضوعات كل مجال

د. الزمن

2. التصميم للمعالجة ودمج أدوات المعالجة

(المدونات اللغوية النصية وحدها ليست من ضمن لغويات/لسانيات المدونات):

<https://bog.gov.sa/ScientificContent/JudicialBlogs/Pages/default.aspx>

# اللسانيات ومعالجة اللغة الطبيعية

## • اللسانيات (معالجة اللغة الطبيعية)

1. الخلفية اللغوية

2. الخلفية الحوسبيّة والإحصائية

## • معالجة اللغة الطبيعية (اللسانيات)

1. المعالجة القبلية: التنظيف/ التوسيم/ التحشية

2. المعالجة البعدية: التقييم اللغوي والتقويم الآلي

3. المعالجة الإحصائية: الإحصاءات الوصفية/ الإحصاءات الاستدلالية/ النماذج الإحصائية

## • العلاقة بين اللسانيات ومعالجة اللغة الطبيعية

1. من اللسانيات: الفرضيات والنظريات والتقييم اللغوي

2. من حوسبة اللغة: اختبار الفرضيات والنظريات والتقييم الآلي (أداء النماذج)

# المناهج المدونية اللسانية الإحصائية

- هناك أدوات وهناك مكاتب في البايثون لتطبيقات هذه المناهج

1. التكرار
2. التصاحب
3. الكشاف السياقي
4. الأنماط (التكرار والتصاحب والكشاف)
5. النظم (التطريز) الدلالي
6. المجموعة الدلالية أو التفضيلات الدلالية
7. قائمة الألفاظ
8. سحابة الكلمات

# مناهج معالجة اللغة الطبيعية

- معالجة اللغة الطبيعية منهج من مناهج تعلم الآلة

- مناهج معالجة اللغة الطبيعية

1. مناهج القواعد rules based

2. مناهج الإحصاء statistics

3. الشبكات العصبية neural networks

- أساليب تعلم الآلة:

1. الموجه: (تدريب النموذج بالاعتماد على الإجابات الصحيحة)

2. غير الموجه: (التصنيف والتجميع والجوار واكتشاف النماذج الخفية)

3. المعزز: توليد بيانات من ذاته، وتعلم مخرجاتها ونتائجها، لتحقيق فهم أعمق

# معالجة اللغة الطبيعية

## • التطبيقات:

1. الكلمات الفعلية tokens
2. التجزئة حدود الفواصل بين أنواع الكلمات، أو بين الجمل، أو بين الفقرات
3. المتتابعات اللفظية n-grams
4. الموسمات النحوية POS taggers / التحليل النحوي parsing والإعراب الاعتمادي
5. كشف الكيانات والعبارات الاسمية

## • الخوارزميات:

1. نمذجة اللغة (تعلم الخصائص)
3. نماذج ماركوف الخفي التوليدية generative (نمذجة الخصائص البسيطة)
4. نماذج الحقول العشوائية المشروطة التمييزية discriminative (نمذجة الخصائص المعقدة والمتداخلة، مثل: الموسمات النحوية)

## • مخرجات معارف معالجة اللغة الطبيعية:

حدود الكلمة/ حدود الجمل/ الكيانات/ الموضوعات المفتاحية في النصوص/ تلخيص النصوص/ السؤال والجواب/ تحليل المشاعر

# الختام: اتجاهات المشاريع

- المشاريع اللسانية: تطبيقات أدوات ومكتبات معالجة اللغة الطبيعية.
- المشاريع الإنسانية: رقمنة تخصصات العلوم الإنسانية والاجتماعية > تخزين المعارف وتحليلها إدارتها.
- المشاريع الحاسوبية: بناء النماذج > تقييم النماذج > تحسين النماذج.