

المحاضرة 11

مهارات الحاسوب ومعالجة النصوص

المرتكزات التقنية والمعرفية لطلبة علم النفس التربوي

تمثل مهارات الحاسوب ومعالجة النصوص حجر الزاوية في التكوين المعاصر لطلبة علم النفس التربوي، حيث لم يعد الحاسوب مجرد أداة مكتبية ثانوية، بل صار امتداداً للعمليات المعرفية والبحثية التي يمارسها الأخصائي في الميدان التعليمي. إن الانتقال من العصر التقليدي إلى العصر الرقمي فرض تحولاً في السوسيو سكوب أو المنظار الاجتماعي الذي نراقب من خلاله السلوك البشري، مما أدى إلى ظهور ما يعرف بالعلوم الاجتماعية الحاسوبية التي تستفيد من القوى الحاسوبية الهائلة لفهم التفاعلات الإنسانية المعقدة.¹ (محتوى الدروس من 1 إلى 10 كما ورد سابقاً...)

الدرس الحادي عشر: توحيد النصوص في بايثون (Text Unification)

في هذا الدرس، سنتناول عملية توحيد النصوص ((Unification)، وهي مرحلة جوهرية في معالجة اللغات الطبيعية (NLP) تهدف إلى تقليل التباين الإملائي في النصوص لضمان أن الكلمات المتشابهة دلاليًا تُعامل ككيان واحد برمجيًا.³ بالنسبة لباحث علم النفس التربوي، هذه العملية هي التي تضمن ألا تضيع بيانات الطلاب بسبب أخطاء إملائية بسيطة.

1. لماذا نحتاج لتوحيد النصوص العربية؟

اللغة العربية لغة غنية بتبايناتها الإملائية، وغالباً ما يكتب المستخدمون (سواء كانوا معلمين أو طلاباً) الكلمات بطرق مختلفة قد لا يراها الإنسان مشكلة، لكنها تربك محركات البحث وخوارزميات التحليل الإحصائي.⁴

- مثال: الكلمات (أحمد، احمد، أحمد) تشير لنفس الشخص، لكن بايثون يراها 3 سلاسل نصية مختلفة.
- الهدف: دمج هذه الأشكال في صورة نمطية واحدة (احد) لتسهيل عملية العد والإحصاء.

2. أهم عمليات التوحيد (Unification Pipeline)

تتضمن عملية التوحيد عدة خطوات تقنية نقوم بتنفيذها على النصوص العربية الخام:

1. توحيد الألف (Alif Unification): تحويل كافة أشكال الألف (أ، إ، آ) إلى الألف العادية (ا).
2. توحيد الناء المربوطة والهاء (Ta-Marbuta Unification): في كثير من الأحيان يكتب الناس مدرسه بدلاً من مدرسة، والتوحيد يعالج هذا اللبس بجعلها جميعاً تنتهي بحرف واحد (عادة الهاء أو

```
import pyarabic.araby as araby

text = أحمد ذهب إلى المدرسة
# 1. إزالة الحركات
clean_text = araby.strip_tashkeel(text)
# 2. توحيد الألف
unified_text = araby.normalize_alef(clean_text)

print(unified_text) # المخرجات: احمد ذهب الى المدرسة
```

5. الفرق بين التوحيد (Unification) والتطبيع (Normalization)

رغم تداخلهما، إلا أن هناك فرقاً دقيقاً يهتم الباحث:³

- **التوحيد:** يركز على الرموز والحروف المتشابهة شكلاً (أ، إ، آ → ا).
- **التطبيع (المحاضرة القادمة):** عملية أوسع تشمل تنظيف النص من الروابط، الإيموجي، الرموز الخاصة، وتحويل النصوص إلى صيغ معيارية (مثل Unicode Normalization).

6. الأثر على جودة البحوث التربوية

إن إهمال مرحلة التوحيد قد يؤدي إلى نتائج مضللة في البحث العلمي:

- **خطأ في التكرارات:** قد تظهر كلمة قلق بتكرار منخفض لأن الباحث لم يوحد القلق وقلق وبالقلق.¹¹
- **صعوبة البحث:** عدم القدرة على استرجاع سجلات الطلاب الذين كُتبت أسماؤهم بدون همزات.
- **ضعف النمذجة:** خوارزميات الذكاء الاصطناعي ستتعامل مع الكلمة الواحدة كعدة مفاهيم، مما يضعف قدرتها على التنبؤ بسلوك الطالب.⁹

الخلاصة

يا أبنائي، إن توحيد النصوص هو عملية فلترة تجعل البيانات الخام صالحة للعلم.³ من خلال هذه المهارة، نضمن أن لغة الأرقام تعبر بصدق عن لغة الحروف، وأن كل أحمد في سجلاتكم هو ذاته أحمد في تحليلاتكم الإحصائية.

في الدرس القادم (رقم 12)، سنختم هذه السلسلة بموضوع **تطبيع النصوص ((Normalization**، حيث سنتعلم اللمسات النهائية لتحويل النص إلى مادة معيارية جاهزة لأرقى أنواع التحليل الإحصائي والذكاء الاصطناعي.³

■ تمرين عملي:

اكتب قائمة بـ 5 أسماء لزملائك مع تعمد كتابة بعضها بالهمزة وبعضها بدونها، ثم حاول تخيل كيف ستقوم دالة `normalize_alef` بتحويلها جميعاً إلى شكل واحد.