

✍️ المحاضرة 9: استخراج المعلومات (Information Extraction) - الجزء الأول

في هذه المحاضرة، ننتقل إلى مستوى أعمق من التعامل مع النصوص. إذا كانت الدروس السابقة قد علمتنا كيف "نخزن" البيانات، فإن هذا الدرس يعلمنا كيف "نستخلص" المعاني والكيانات من داخل النصوص غير المنظمة (مثل المقابلات الإكلينيكية والتقارير المدرسية) لتحويلها إلى بيانات منظمة يمكن قياسها إحصائياً.

🌸 تحويل الفوضى إلى نظام: (Unstructured vs Structured)

يعمل الأخصائي النفسي عادة مع نصوص "غير مهيكلة" (فقرات طويلة، ملاحظات عفوية). تكمن عبقرية البرمجة هنا في قدرتها على تحويل هذه الفقرات إلى جداول منظمة تحتوي على:

- **الكيانات:** (أسماء الطلاب، أسماء المدارس).
- **العلاقات:** (نوع الاضطراب السلوكي المرتبط بكل طالب).
- **الأحداث:** (تاريخ وقوع نوبة الغضب أو التئمر).

👤 التعرف على الكيانات المسماة (NER): المساعد الرقمي الذكي

تعد تقنية **Named Entity Recognition (NER)** واحدة من أقوى أدوات الذكاء الاصطناعي في معالجة اللغات الطبيعية. هي عملية آلية تهدف لرصد وتصنيف الكلمات الهامة داخل النص:

- **الأشخاص:** التعرف آلياً على أسماء الطلاب أو المعلمين في تقرير ضخم.
- **المؤسسات:** رصد أسماء المدارس أو مراكز إعادة التأهيل.
- **المواقع:** تحديد التوزيع الجغرافي لظواهر تربوية معينة (مثل مناطق انتشار صعوبات التعلم).
- **المكتبات المستخدمة:** تعتمد في بايثون على مكتبات عملاقة مثل **spaCy** و **NLTK** التي توفر نماذج مدربة قادرة على "فهم" سياق الكلمات.

😊 تحليل المشاعر (Sentiment Analysis): قراءة ما بين السطور

كأخصائيين، يهمنا جداً رصد "النبرة الانفعالية" في كتابات الطلاب أو المعلمين. يساعدنا تحليل المشاعر في تصنيف النصوص إلى (إيجابية، سلبية، محايدة) بناءً على الكلمات المفتاحية:

- **مثال تطبيقي:** إذا كتب طالب "أشعر بإحباط شديد"، يقوم البرنامج فوراً برصد كلمة "إحباط" وتصنيف الحالة ضمن "المشاعر السلبية"، مما يعطي إنذاراً مبكراً للأخصائي للتدخل.

🔵 ميثاق الشرف الرقمي: أخلاقيات استخراج البيانات

عندما نستخدم الآلة لاستخراج الأسماء والتشخيصات، نصبح أمام مسؤولية أخلاقية كبرى تتلخص في ثلاثة مبادئ:

1. **إخفاء الهوية (Anonymization):** قبل معالجة البيانات ونشرها بحثياً، يجب استبدال الأسماء الحقيقية برموز (الطالب أ، الطالب ب) لحماية الخصوصية.
2. **الموافقة المستنيرة:** يجب إبلاغ المشاركين أو أولياء أمورهم بأن بياناتهم ستعالج آلياً لأغراض بحثية وتطويرية.
3. **تجنب التحيز الآلي:** الخوارزميات قد تكون منحازة أحياناً؛ لذا يجب أن يظل القرار النهائي بيد الأخصائي البشري، والآلة هي مجرد "مساعد" يقدم الاحتمالات.

🚀 الخلاصة: الباحث كمنقب عن المعرفة

لقد تعلمنا اليوم أن النص ليس مجرد مجموعة كلمات، بل هو "منجم" للمعلومات. باستخدام تقنيات IE و NER، يستطيع الباحث التربوي:

- رصد الأنماط السلوكية في وقت قياسي.
- تحويل مئات الساعات من المقابلات الصوتية المفرغة إلى جداول إحصائية دقيقة.
- المساهمة في بناء "أنظمة إنذار مبكر" للمشكلات النفسية داخل المدارس.

💡 **رسالة المحاضرة:** إن الحفر داخل النصوص لاستخراج الكنوز المعلوماتية هو ما يصنع الفارق بين الباحث التقليدي والباحث الرقمي المحترف. في الدرس القادم، سنتعمق أكثر في كيفية تطبيق هذه التقنيات عملياً باستخدام مكتبة spaCy.