المحاضرة الثامنة : معالجة النصوص والبيانات

تهدف هذه المحاضرة الى تحقيق الاهداف التالية :

- **ان يتعرف الطالب على العمليات الاساسية في البرمجة**
- **ان تعلم صياغة العمليات بلغة الدوال**

معالجة النصوص(String Processing)

معالجة النصوص هي عمليات تحرير، تحليل، وتعديل السلاسل النصية مثل التقطيع، البحث، الاستبدال، وتغيير حالة الأحرف.[String Processing Guide, Real Python 2025]

عمليات أساسية:

العملية	الوصف	مثال(Python)
التقطيع(Split)	فصل النص إلى كلمات	words = "Hello World".split()⟶['Hello', 'World']
البحث(Find)	إيجاد موقع نص	"Python"[2:5]⟶'tho'
الاستبدال(Replace)	تغيير كلمات	"Hello".replace("H", "J")⟶'Jello'
تغيير الحالة	upper/lower	"Python".upper()⟶'PYTHON'
الطول	عدد الأحرف	len("Python")⟶6[Real Python]

معالجة البيانات(Data Processing)

تحويل البيانات الخام إلى معلومات مفيدة باستخدام هياكل البيانات والعمليات الحسابية [Data Processing Handbook, Pandas 2025].

هياكل البيانات الشائعة:

```python
# قوائم (Lists)
fruits = ["تفاح", "موز", "برتقال"]
fruits.append("مانجو")  # إضافة

# قواميس (Dictionaries) - تخزين بيانات مترابطة
student = {
    "name": "أحمد",
    "grade": 95,
    "subjects": ["رياضيات", "علوم"]
}

# قراءة CSV بـ Pandas
import pandas as pd
df = pd.read_csv("data.csv")
print(df.head())  # أول 5 صفوف [Pandas documentation]
```

أمثلة عملية متكاملة

أ. تحليل نص:

```python
text = "Python is great for data processing"
words = text.lower().split()
word_count = {word: words.count(word) for word in words}
print(word_count)
# {'python': 1, 'is': 1, 'great': 1, 'for': 1, 'data': 1, 'processing': 1}
```

ب. معالجة بيانات الطلاب:

```python
students = [
    {"name": "محمد", "grade": 87},
    {"name": "فاطمة", "grade": 92},
]

# حساب المتوسط
avg = sum(s["grade"] for s in students) / len(students)
print(f"المتوسط: {avg}")  # 89.5 [web:412]
```

4.مكتبات متخصصة

المكتبة	الاستخدام	مثال
Pandas	معالجة جداول البيانات	df.groupby('city')['sales'].sum()
NumPy	حسابات رياضية سريعة	np.mean([1,2,3,4])→2.5
Regular Expressions (re)	بحث نصوص معقد	re.findall(r'\d+', "abc123")→['123']
BeautifulSoup	استخراج منHTML	تحليل صفحات الويب[Scrapy Docs]

.معالجة النصوص(Text Processing)

تشير معالجة النصوص إلى جميع العمليات التي يتم إجراؤها على البيانات النصية (مثل الجمل، الفقرات، الوثائق) لتهيئتها وتحليلها. وتُستخدم بشكل واسع في تطبيقات **معالجة اللغة الطبيعية.(NLP)**

المفهوم	الوصف والهدف	الأهمية للباحث
التجزئة (Tokenization)	تقسيم النص إلى وحدات أصغر تُسمى "رموز" (Tokens)، قد تكون كلمات، حروف، أو جمل.	هي الخطوة الأولى لفصل العناصر اللغوية قبل عدّها أو تحليلها.
التنظيف (Cleaning)	إزالة العناصر غير المرغوب فيها من النص، مثل علامات الترقيم، والأرقام، والرموز الخاصة، وتوحيد حالة الأحرف.	ضروري لتقليل الضوضاء في البيانات وضمان دقة التحليل الإحصائي للكلمات.
إزالة الكلمات الموقوفة (Stop Word Removal)	حذف الكلمات الشائعة التي لا تحمل قيمة دلالية كبيرة (مثل: "من"، "في"، "الـ...").	يساعد في تركيز التحليل على الكلمات ذات الأهمية المعجمية (Content Words).
التجذيع (Stemming & Lemmatization)	تقليص الكلمات إلى جذرها الأساسي أو صيغتها المعجمية.	يضمن عدّ الكلمات ذات الأصل الواحد مرة واحدة (مثل: "يكتبون"، "كتابة"، "كاتب" تُرد إلى "كتب").

معالجة البيانات(Data Processing)

تشير معالجة البيانات إلى مجموعة الخطوات المنظمة التي يتم فيها تحويل البيانات الخام إلى معلومات ذات مغزى وقابلة للاستخدام. وتتم هذه العملية غالباً باستخدام برامج **جداول البيانات**)مثل (Excel أو مكتبات برمجية (مثل Pandas في.(Python

أ. المراحل الأساسية(Basic Phases)

1. **الجمع :(Collection)** إدخال البيانات من مصادر مختلفة (قواعد بيانات، استبيانات، استخلاص من الويب.(
2. **التجهيزوالتنظيف :(Preparation & Cleaning)** إصلاح الأخطاء، التعامل مع القيم المفقودة (Missing Values)، إزالة البيانات المكررة .**(هذه هي المرحلة الأكثر استهلاكاً للوقت)**.
1. **المعالجة :(Processing)** تطبيق خوارزميات أو عمليات حسابية لتحويل البيانات (مثل حساب المتوسطات، تجميع البيانات، أو إنشاء متغيرات جديدة.(
2. **التخزين :(Storage)** حفظ البيانات المعالجة في صيغة يسهل الوصول إليها واستخدامها لاحقاً.
1. **التحليل والتفسير :(Analysis & Interpretation)** استخدام الإحصاءات أو التعلم الآلي لاستخلاص الأنماط والنتائج وتفسيرها.

ب. أدوات شائعة لمعالجة البيانات:

الأداة	الوظيفة الرئيسية	المثال في البرمجة
الفرزوالتصفية (Sorting & Filtering)	**تنظيم البيانات حسب قيمة معينة، أوعرض مجموعة فرعية منها فقط (لتحليل عينة محددة.(**	**DataFrame.sort_values(), DataFrame[condition] (في Pandas)**
التجميع (Aggregation)	**تلخيص البيانات في مجموعات (مثل حساب متوسط درجات الطلاب حسب التخصص.(**	**DataFrame.groupby().mean()**
دمج البيانات (Data Merging)	**ضم مجموعتين من البيانات معاً بناءً على مفتاح مشترك (مثل ربط سجلات الطلاب ببيانات حضورهم.(**	**pd.merge(df1, df2)**

الأهمية المنهجية: تضمن هذه العمليات أن تكون البيانات (النصية والكمية (**عالية الجودة وموثوقة**)قبل بدء التحليل المنهجي، مما يؤدي إلى استنتاجات بحثية أكثر دقة.(Lynch, 2020)