

XML i XSD: Validació i Conversió

INDEX

INTRODUCCIÓ.....	3
1. Cerca i anàlisi.....	4
1.1 Per què hem triat aquest lloc?.....	4
1.3 Detecció d'un XSD.....	6
2. Validació.....	8
3. Conversió.....	9
3.1 Conversió a JSON.....	9
4. Conclusió.....	11

INTRODUCCIÓ

Aquesta pràctica té com a objectiu principal conèixer com s'utilitza l'XML en webs públiques i entorns reals.

Primer, cal localitzar i descarregar un fitxer XML accessible públicament.

Un cop descarregat, s'ha d'identificar si aquest fitxer està vinculat a un esquema XSD que defineixi la seva estructura. Si existeix, es realitzarà la validació de l'XML amb eines online o locals. En cas contrari, es crearà un XSD senzill per validar-lo.

També es treballarà la conversió del document XML a altres formats com JSON, CSV o HTML. Aquesta conversió permet entendre com es poden reutilitzar les dades. A més, es valorarà la importància del XML com a eina d'intercanvi d'informació entre aplicacions. S'utilitzaran eines bàsiques com navegadors, editors de text i validadors online. Finalment, es farà una reflexió sobre l'experiència.

Aquesta pràctica té com a finalitat introduir-nos en l'ús real d'XML com a eina d'intercanvi d'informació a través de webs públiques, així com explorar la seva validació i conversió a altres formats.

1. Cerca i anàlisi

S'han utilitzat les dades meteorològiques proporcionades per l'AEMET a través del seu portal de dades obertes: [AEMET](#)

Una forma d'accedir a les dades és amb la clau API i un cop obtinguda la clau, es poden descarregar arxius XML amb prediccions meteorològiques detallades per municipi. Per exemple, la predicció per a Mollet del Vallés.

1.1 Per què hem triat aquest lloc?

Perquè és fàcil d'accedir, els XML són estructurats i varien en temàtiques que poden ser útils per a diferents tipus d'anàlisis. A més, és un portal oficial.

1.2 Anàlisi de l'estructura

Els fitxers XML proporcionats per l'AEMET tenen una estructura jeràrquica que inclou informació detallada sobre la predicció meteorològica.

Podem trobar el fitxer XML de la web seguint el següent enllaç:

https://www.aemet.es/xml/municipios/localidad_08124.xml

Aquesta estructura jeràrquica facilita la comprensió i processament dels documents legislatius.

A continuació, per poder entendre l'estructura tenim la següent explicació:

En la primera línia trobem aquestes característiques:

- `id="08124"`: Codi INE de Mollet del Vallès.
- `version="1.0"`: Versió de l'estructura XML.
- `xsi:noNamespaceSchemaLocation`: URL on es troba l'esquema XSD que valida l'XML.

Troblem també les etiquetes de nom, província i elaborado les quals ens indiquen la localitat a la qual fa referència la predicció, la província Barcelona i la data i hora en què s'ha generat la predicció.

A més, cada element `<dia>` conté la predicció per a una data concreta. Dins de cada dia, hi ha informacions dividides en períodes (ex. `00-06`, `06-12`, etc.).

També tenim les etiquetes de `<prob_precipitacion>`, `<cota_nieve_prov>`, `<estado_cielo>` i més que ens parlen a sobre del vent i la temperatura. Aquestes etiquetes nombrades, ens indiquen el percentatge de probabilitat que plogui en aquest interval horari, la altitud (en metres) a partir de la qual es preveu que la precipitació sigui en forma de neu entre d'altres.

En conclusió, aquest document XML de l'AEMET ofereix:

- Prediccions meteorològiques completes **per a diversos dies**.
- Informació dividida per **franges horàries**.
- Dades sobre **pluja, neu, núvols, vent, temperatures, humitat i UV**.

És ideal per a consultes automàtiques, aplicacions mòbils, webs o sistemes de predicció.

1.3 Detecció d'un XSD

L'arxiu XML definit per aquest esquema representa una predicció meteorològica d'un municipi. L'element principal és `<root>`, que inclou informació general i una predicció per dies.

Els primers elements són:

- `<origen>` (opcional): pot indicar qui ha generat les dades.
- `<elaborado>`: data i hora de creació de la predicció (format ISO: AAAA-MM-DDTHH:MM:SS).
- `<nombre>` i `<provincia>`: identifiquen el municipi i la seva ubicació.

Atributs de `<root>`:

- `id`: identificador del municipi.
- `version`: versió del document.
- `xsi:noNamespaceSchemaLocation` (opcional): pot ajudar a validar el document amb el seu esquema.

La secció més important és `<prediccion>`, que conté una llista de `<dia>`, un per cada dia de la predicció. Cada `<dia>` té l'atribut `fecha` i una gran varietat de dades meteorològiques:

- `<prob_precipitacion>`: indica la probabilitat de pluja per franges horàries (`periodo`).
- `<cota_nieve_prov>`: altitud mínima on es pot esperar neu, també per hores.
- `<estado_cielo>`: descriu l'aspecte del cel (clar, núvol, boira...), amb atributs `periodo` i `descripcion`.
- `<viento>`: conté direcció i velocitat del vent, amb un atribut `periodo`.
- `<racha_max>`: mostra les ràfegues de vent més fortes (si n'hi ha).
- `<temperatura>`: inclou màxima, mínima i valors per hores (`dato hora="..."`).

- <sens_termica>: com es percep la temperatura realment (amb la mateixa estructura que la temperatura).
- <humedad_relativa>: màxima, mínima i valors horaris d'humitat.
- <uv_max> (opcional): índex ultraviolat màxim del dia.

2. Validació

S'ha utilitzat l'eina xmlint per validar un fitxer XML del AEMET.

Seguirem l'enllaç i al entrar veurem una capsa on ens indica que hem d'escriure allà el nostre element XML per tal de validar-ho.

Primer de tot hem de descarregar el fitxer XML i en el cas d'haver-hi XSD fer la validació.

El document XML l'hem descarregat via terminal amb la comanda:

```
wget https://www.aemet.es/xml/municipios/localidad_08124.xml
```

→<https://xmllint.com/>



Figura 2. Validació

La validació ha estat satisfactòria, confirmant que el fitxer XML compleix amb l'esquema establert.

3. Conversió

En aquesta fase, s'ha dut a terme la transformació del fitxer XML a altres formats més accessibles i fàcils de manipular, com són **JSON**. Aquest procés és fonamental per facilitar la integració del contingut XML en aplicacions web, eines d'anàlisi de dades o visualitzadors.

3.1 Conversió a JSON

Per a la conversió de l'XML a format **JSON**, s'ha utilitzat la llibreria **xmldict** de Python. Aquesta llibreria permet transformar estructures XML en diccionaris Python, que posteriorment es poden exportar fàcilment com a fitxers JSON. Adjuntem la conversió a la entrega de la pràctica..

El fitxer JSON conté informació meteorològica detallada del municipi de **Mollet del Vallès**. La primera part és la secció "origen", que ens diu qui ha generat les dades ("productor": "AEMET"), l'empresa responsable ("elaboradoPor"), i també ens proporciona un enllaç ("enlace") a la font oficial. A més, hi ha una "notaLegal" amb informació sobre l'ús públic i els drets del contingut. L'etiqueta "elaborado" indica la data i hora en què s'ha creat la predicció.

A continuació, trobem "prediccion", que conté una llista de dies. Cada dia es representa com un objecte dins l'array "dia", i inclou molta informació:

- **"fecha"**: el dia de la predicció (format AAAA-MM-DD).
- **"probPrecipitacion"**: un array amb la probabilitat de pluja per franges horàries (00-06, 06-12, etc.).
- **"estadoCielo"**: també dividit per hores, indica si estarà clar, núvol, amb boira, etc. Cada valor té un "descripcion".
- **"viento"**: mostra la direcció ("direccion") i la velocitat ("velocidad") del vent per franges horàries.
- **"rachaMax"**: si hi ha pics de vent forts, aquí s'indica la ràfega màxima.
- **"temperatura"**: inclou la "maxima" i la "minima" del dia, i un array "dato" amb temperatures per hora.
- **"sensTermica"**: com es percep la temperatura real, amb "maxima", "minima" i valors per franges horàries.

- **"humedadRelativa"**: mostra la "maxima" i "minima" del dia, i també valors per franges horàries.
- **"uvMax"**: és l'índex ultraviolat màxim, que indica el risc solar.

A més, hi ha una etiqueta "cotaNieveProv" que indica a quina altitud es pot esperar neu a la província. També poden aparèixer "estadoSuelo" o "probNieve" en altres dies si la situació ho requereix.

Finalment, fora de "prediccion", hi ha "nombre", "provincia", "altitud", "id" i "version", que identifiquen el municipi i l'altura a què es troba.

-<https://pypi.org/project/xmltodict/>

4. Conclusió

Primer de tot, hem utilitzat dades de l'Agència Estatal de Meteorologia (AEMET), que ofereix informació oberta i accessible per a tothom. Els documents estan en format XML, que és un format ordenat i estàndard, ideal per poder-los utilitzar en moltes aplicacions diferents.

L'XML que hem analitzat està ben estructurat, amb parts clares com `<prediccion>`, `<dia>`, `<temperatura>`, o `<estado_cielo>`. Aquestes parts ajuden a entendre fàcilment la informació important, com la data, la temperatura o com estarà el cel. Així, tant les persones com els programes poden llegir-ho sense problemes.

A més, hem comprovat que l'arxiu XML és correcte perquè hem fet servir un esquema XSD personalitzat. Aquesta validació ens assegura que el document segueix les normes i està ben fet, sense errors.

Finalment, hem pogut convertir aquest XML a format JSON amb la llibreria `xmltodict` de Python. Aquesta conversió mostra que el format XML és molt flexible i es pot adaptar bé per treballar amb eines i programes moderns.