**Első nap**

Az ETL ciklus lépésein belüli feladatok

1. Cycle initiation
2. Build Reference Data
3. Extraction (from sources)
4. Validation
5. Transformation
6. Stage
7. Audit reports
8. Publish
9. Archive

Ha kérdés van egy fejlesztendő projekt kapcsán, akkor az kerüljön be a specifikációba, majd kerüljön megválaszolásra.

Ha a specifikációban egy p mondat nem egyértelmű, akkor az kerüljön részletezésre.

(

Van kérdés egy programmal kapcsolatban?

Valami nem egyértelmű a specifikációban?

A specifikációban minden kérdés megválaszolásra került?

)

Ha a válasz igen az előző három kérdésre, akkor kész a specifikáció

Konklúzió:

Nem történt egy olyan projekt definiálása, amit finomítva eljuthattam volna egy szkripthez, ami elvégzi az adott munkát.

* -

S(x1, x2, …, xn) egy függvény, aminek az értéke egy releváns mondat egy data extraction munkára vonatkozóan.

(x1, x2, …, xn) szám n-es leírja egy data extraction munka paramétereit.

Mikor nyújt S(x1, x2, …, xn) bizonyítást arra vonatkozóan, hogy képes vagyok elvégezni a megpályázott munkát?

* Akkor, ha S(x1, x2, …, xn)-t kiértékelve a megrendelő arra a kérdésre, hogy "megfelel-e az aktuálisan vizsgált pályázó a munkára" igennel válaszol.

Az általam biztosított struktúra által a munkának elvégezhetőnek kell lennie.

* -

Use Case-ek definíciója munkatípusokra vonatkozóan.

Steps

1. Data Extraction
2. Data transformation



1. Data Loading



* -

A kódolási képességem még mindig nem a legjobb, rengetegen sokkal jobban kódolnak nálam.

* Solve problems related to tasks included in the ETL process within Python



-o-

**2. NAP**



-o-

**4. NAP**

****

-o-

**5.NAP**

Header:

header= {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/96.0.4664.45 Safari/537.36',

'Accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,\*/\*;q=0.8',

'Accept-Charset': 'ISO 10646, utf-8',

'Accept-Encoding': '\*',

'From':'solteszistvan@outlook.com'}

Question:

* How to scrape tables?

**6.NAP**

**Parameters of a Data Extraction work:**

* Source
  + Format
* Output
  + Format
  + Data Structure
* Constraints
  + What makes a data valid within the valid?
* Logging
  + Formulas Used
  + Process used
  + System used

**7.NAP**

-> ?

**Facebook project**

Data structure for the facebook project:



ID- Data pairs for the facebook project



-o-

Mapping graphs to code and rewriting ER diagrams



-o-

**8.NAP**

Password code:



Cipher.py:



-o-

**9.NAP**

Interesting categories:

* Category of pictures
  + What is the definition of a picture?
* Category of Graphs
* Category of Computer Code

-o-

Learning

* Categories
* Generating Files
  + Generating Source Code
    - Python
  + Generating Word files

Requests

Some basic requests and methods



Session object





**A heuristic process**

1. Generate picture

2. Make distinctions between objects in a picture

3. Put those objects into sets

4. Describe those objects

-o-

**DAY 11**

Defining small problems that can be solved using creativity.

**Example**:

Retrieve a form from a valid HTML network related to a POST request, involving Facebook access.

Validness is undefined

Format of the solution: Python dictionary

* (Definition, Constaints, output)

-o-

**DAY12**

Először is:

Nem tudok gráfokat tárolni, ami szomorú.

Kell egy osztály, ami meg tud nekem valósítani egy adatszerkezetet, amiben gráfokat tárolok.

Ez fontos.

-o-



-o-



-o-



-o-

**DAY13**





Regex can be used to find patterns within unclean data, to locate unclean data.

1. Define pattern
2. Define an iterative way to parse an object
3. For each iteration, use the pattern to find unclean data
4. If unclean data is found, clean it.

Cleaning data

PROBLEMS

Convert date to python date objects

Storing numeric objects to integers

replace strings to numerical values if their value are equal.

Standardizing data

Mapping intervals to certain values

Grouping objects

- -

Define clean data

-o-

Questions:

* What is a python datetime object?

**DAY14**



-o-

A Challange-el késtem egy hetet, nem tudtam időben befejezni.

Nem volt egy megfelelő modellem, amit finomítva kaphattam volna egy projektet, ami elkészül.

Projekt definíció hiányában ez várható volt.

A challange-et megismétlem.

Ami fontos, hogy megfelelő projektdefinícióval rendelkezzek.

Hasznos lenne, ha tudnék gráfokat tárolni Python kódon belül.

Hasznos lenne, ha rendelkeznék megfelelő szerkezettel a Python kódom formálására vonatkozóan.

* Ehhez modulokat, megfelelő könyvtár hierarchiát, valamint modulokra vonatkozó formalizmusokat tudok használni.

-o-

Nem lenne rossz az ostoba ötletek mértékét csökkenteni.

Mi az ostobaság mértéke?