

Zuzana Nevěřilová
Hana Žižková
2024/25

CORE147

DIGITÁLNÍ DATA
V HUMANITNÍCH
A SOCIÁLNÍCH



APLIKAČNÍ PROGRAMOVÉ ROZHRANÍ (API)

Příklad digitalizace z minulé přednášky

1. Registrace v Google Cloud
2. Aktivace Google Vision
3. Získání klíče (případně platba)
4. Použití balíčku v Pythonu pro Google Vision
5. Použití klíče
6. Vyvolání funkce OCR z Google Vision
7. Získání OCR dokumentu
 - Text
 - Souřadnice

P R E F A C E.

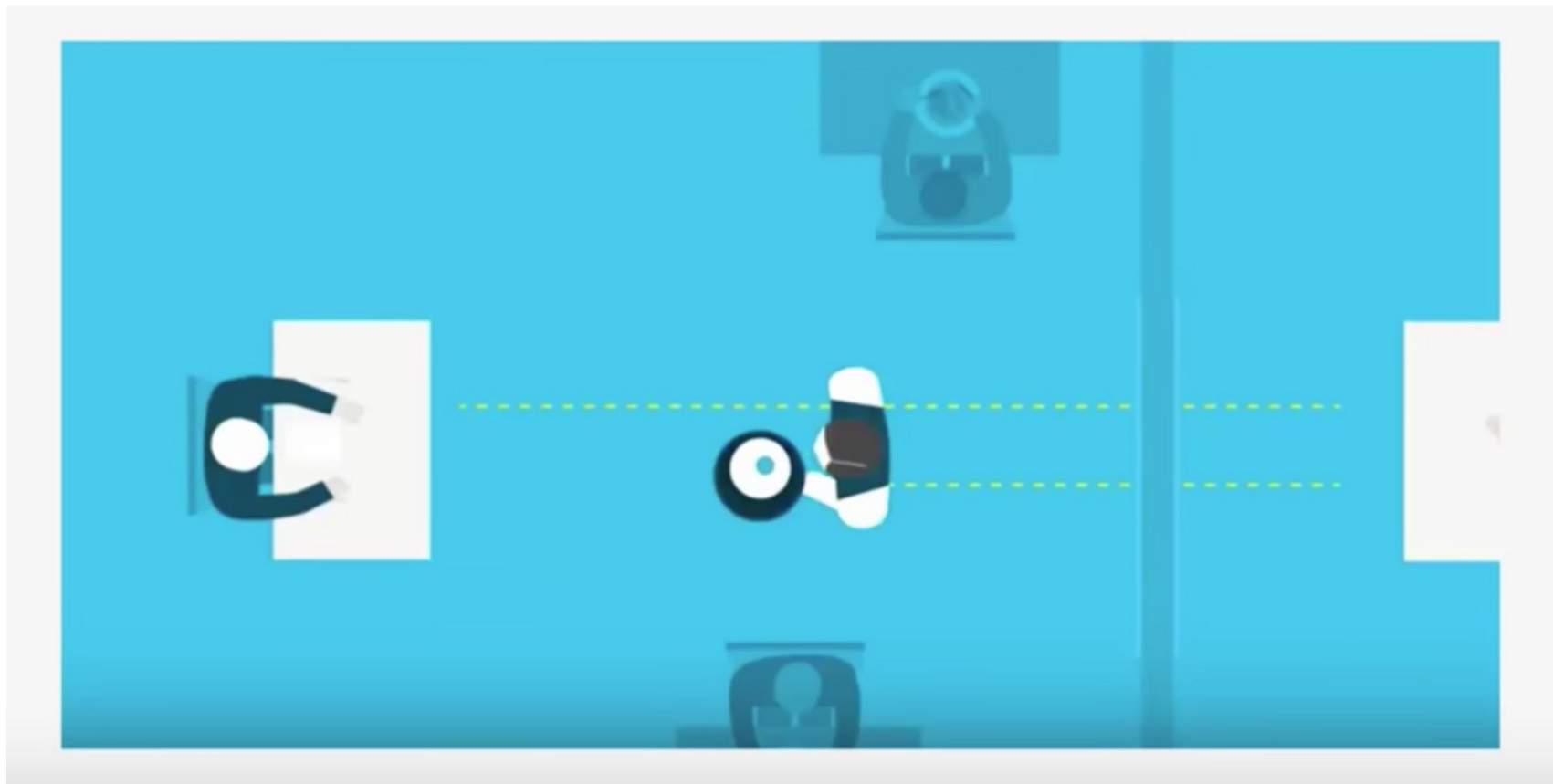
THE editor can now congratulate his readers and himself, on having brought this work to a conclusion. He expected to have done so, at a much more early date. Unforeseen difficulties arose to prevent the accomplishment of his design*. The best apology for this delay, is to point out some of the principal improvements that it was found requisite to make to the original publication, and which have principally occasioned the tardiness of the editor in fulfilling his engagements.

In the article of *Poland*, besides numerous corrections in point of style, and the omission, or emendation of many exceptionable passages, the history of the establishment of the new constitution in 1791, and its subsequent overthrow, have been newly written for the present edition. The essay on Greek literature, in the succeeding article, has also been added, and another, on a similar subject, under the head of *Turkey in Asia*. With respect to the internal administration of the Ottoman empire, and the domestic manners of its people, some interesting details have been borrowed from Baron de Tott, and M. de Peyssonnel. The article of *Tartary in Asia* will also be found considerably improved. To that of *China* large additions have been subjoined, and in particular, besides other heads, under those of *population and literature*. *India within the Ganges* has, among other additions, an article extending to twelve pages, on the administration of the British East-India Company. The history, under the same head, being eight pages, is, a few paragraphs excepted, entirely new. To *the states of Barbary* there has been added a new article on their history. The united states of America occupy from p. 243, to p. 580, inclusive, in the present volume. It was noticed, in the preface to vol. I. that the account given by Guthrie was altogether trifling. It has been entirely laid aside. Under this head, though the editor has exerted himself to the utmost for obtaining authentic materials, he is sensible that his account is at a very great distance from perfect accuracy. The rev. Jedidiah Morse, author of the *Universal Geography*, has furnished the principal part of it. To render it as complete as possible, application has been made to gentlemen residing in different states; and to their assistance the work is indebted for many interesting particulars that have not hitherto been published.

From the multiplicity of subjects that are to be examined, and the multiplicity of opinions that arise upon every subject, it becomes a task of the greatest delicacy to

* One reason for the delay has been, that the work has extended to fifty-six numbers, instead of forty-eight, which were originally proposed. The eight extra numbers have been delivered gratis to the subscribers, agreeably to engagement.

APLIKAČNÍ PROGRAMOVÉ ROZHRAŇÍ



<http://www.danielgold.net/2019/05/api-lets-make-it-simple.html>

APLIKAČNÍ PROGRAMOVÉ ROZHRAŇÍ

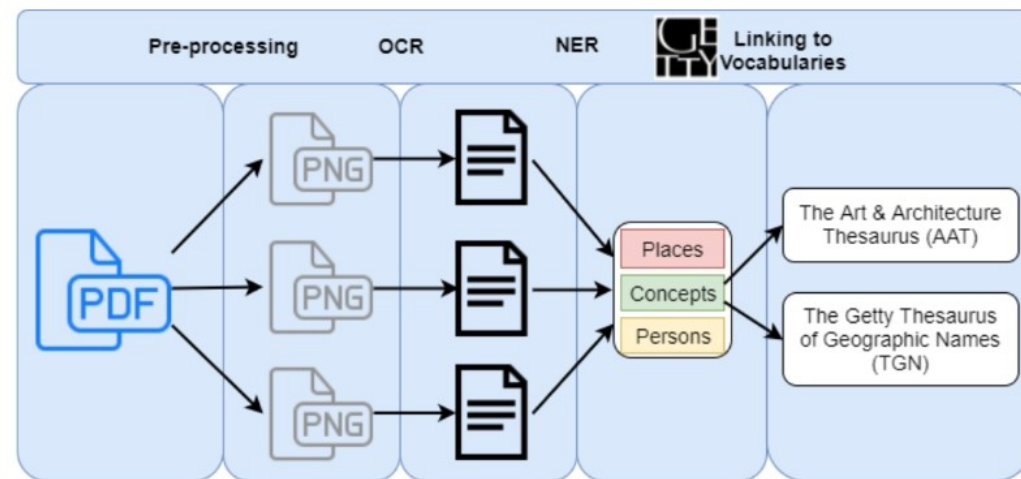
- Restaurace
- Host
- Kuchař
- Číšník
- Objednávka
 - Host si vybere z nabídky
 - Host se umí domluvit s číšníkem
 - Host řekne, co chce
- Vyřízení objednávky
 - Číšník donese jídlo (nebo ne)
- Aplikace
- Uživatel
- Serverová část
- API
- Požadavek (request)
 - Vlastnosti (properties) API
 - Formát (jazyk)
 - Parametry požadavku
- Odpověď (response)
 - Kód odpovědi

APLIKAČNÍ PROGRAMOVÉ ROZHRAŇÍ

- Webová služba, nejčastěji RESTful (Representational State Transfer)
 - Výhoda: požadavek pošleme na URL
 - Nevýhoda: lze pracovat dobře jen s textem
 - Výhoda i nevýhoda: bezstavovost (každý požadavek je nezávislý)
- Balíček/knihovna v programovacím jazyce (např. Python module)
 - Výhoda: uživatel neřeší, kam se posílá jaký požadavek
 - Nevýhoda: integrace do jiné aplikace (v jiném programovacím jazyce)
- Standard pro API: OpenAPI, dříve Swagger
 - Výhoda: vstupy a výstupy jsou standardizované
 - Nevýhoda: ne všichni OpenAPI používají
- API zdarma vs. placená API

POUŽITÍ API V PROJEKTU

1. Rozdělení činností na **malé úlohy**
2. Rozhodnutí, které úlohy jsou obecné
 - OCR
 - Rozdělit text na slova
 - Určit jazyk textu
 - Najít základní tvar slova
 - Najít místo na mapě
 - Najít objekt v obrázku
 - Najít počasí pro danou lokalitu a čas
3. API pro každou úlohu (uvážení ceny, kvality, rychlosti, spolehlivosti, kapacity)
4. Technické požadavky (vstupní a výstupní formáty)
5. Vytvoření **pipeline** pro celou úlohu
 - Konverze formátů podle požadavků API



https://github.com/biktorr/aml_colonialnlp

TYPICKÁ PRÁCE S API

- Co API nabízí?
- Jaké jsou limity?
 - Počet požadavků za minutu, za den, za měsíc
 - Maximální velikost vstupu
 - Maximální velikost výstupu
 - Hard limits vs. Soft limits
- Je třeba klíč (API key, token)?
- Nachystat požadavek
 - Přidat klíč/token
 - Přidat limit
- Odeslat požadavek
- Zkontrolovat typ odpovědi (OK/chyba)
- V případě korektní odpovědi zpracovat výsledek

PŘÍKLAD API: WIKIPEDIA

XML

```
<api batchcomplete="">
  <query>
    <normalized>
      <n from=
        "greek_architecture" to=
        "Greek architecture"/>
    </normalized>
    <pages>
      <page idx= "19493213"
        pageid= "19493213"
        ns="0"
        title= "Greek
          architecture"/>
    </pages>
  </query>
</api>
```

JSON

```
{
  "batchcomplete": "",
  "query": {
    "normalized": [
      {
        "from": "greek_architecture",
        "to": "Greek architecture"
      }
    ],
    "pages": {
      "19493213": {
        "pageid": 19493213,
        "ns": 0,
        "title": "Greek architecture,"
      }
    }
  }
}
```


KDE HLEDAT API

- Zkuste hledat “list of apis” nebo “api marketplace”
 - <https://any-api.com>
 - <https://apilist.fun>
 - <https://rapidapi.com>
 - A mnoho dalších
- Specificky pro jazykové zdroje, využijte infrastruktury CLARIN
- Specificky pro humanitní vědy, využijte infrastruktury DARIAH
- Zjistěte, zda vaše oblíbená webová stránka poskytuje API

DALŠÍ ČTENÍ/SLEDOVÁNÍ

- Abigail Walker: Getting Data for Digital Humanities with APIs: A Gentle Introduction. Digital Humanities@Pratt. 2019.
<https://studentwork.prattsi.org/dh/2019/05/13/getting-data-for-digital-humanities-with-apis/>
- Mark Hall (2021). Introduction to APIs. Version 1.0.0. DARIAH-Campus. [Training module]. https://campus.dariah.eu/id/crdiXpmt0_AkSLZcql6C