

Wissensdatenbank

```
wissensDatenbank.json

[
  {
    "role": "user", "content": "Ich gebe dir ein paar Dialog Beispiele in Vallader damit du die Sprache besser verstehst" },
    {
      "role": "assistant", "content": "Grazia" },
    {
      "role": "user", "content": "Allegra" },
    {
      "role": "assistant", "content": "Allegra, co poss jau gidar" },
    {
      "role": "user", "content": "Be amo tschinch nots e lura n'haja anniversari" },
    {
      "role": "assistant", "content": "Faina üna festina?" },
    {
      "role": "user", "content": "Hier ein kurzer Text damit du besser Vallader kannst. Der Text ist ein Dialog und es geht um Zahnschmerzen: «Auaa», plondscha Lupo, «meis daints fan suul mal.» Lupo schmetta da rusigliar vi da seis öss e til lascha crodar our d'bocca. «Ha! N'ha bain pensà cha teis daints gnian marschs cun magliar glatsch! Uossa hast la merda, uossa varast il prüm dolurs, lura crodarà oura ün daint, lura amo ün e lura amo ün - fin cha tü nun hast plü ingüns. E lura poust be amo magliar schoppa e mösas fin cha tü mourast!», disch Brutus cun tun dad important. «Ojee, ma quai nu lessa!», disch Lupo e guarda tuot sculozzà." },
    {
      "role": "assistant", "content": "Grazia fitg per questa decleraziun" },
    {
      "role": "user", "content": "Hier ein weiterer kurzer Text damit du besser Vallader kannst. Der Text ist ein Dialog und es geht zwei Kinder in einem Stall: Nic e Brutus sun in viadi vers la stalla da sar Men. Nic va minchatant a güdar a sar Men in stalla. El dà lura ün pa fain a las vachas ed a las chavras, pavla la giallinas e quinta ün pa ballas cun sar Men. Sar Men ha eir ün asen ed her es güsta nat ün asenin e quel voul Nic ir a verer. Brutus chi vaiva ün pa lungurus, ha decis dad ir cun el. Nic ha aint stivals verds, ün pêr chotschas da camufladi ed üna giacca da fleece verda e sün seis cheu üna chapütscha cun ala. Uschè va el adüna in gir cur ch'el va a güdar in stalla. «Uei, quia riva meis famagl! Ma chau Nic», cloma sar Men. «Buah, che spüzza chi'd es quia, quai nun es bod na da tgnair oura!», Brutus as tegna il nas. «Ah, e guarda qua, il giat chi ha adüna noscha glüna es eir quia! Che onur! Mo tü parast dad esser ün pa üna gniffa!», sar Men ria. Brutus fa sco sch'el nu dudiss e tegna inavant il nas. Nic ria e dumonda lura a sar Men: «Ingio es l'asenin?» «Gnit cun mai ch'eu as muoss», disch sar Men. Nic, sar Men e Brutus traversan la stalla. Els passan dasper üna pruna vachas via. Quellas maglian, rumaglian e sventuleschan lur cuas per scurrantar las muos-chas. Tuot in ün dandet as doda platsch ... platsch ... platsch - «Bäääh, quellas chajan immez stalla!», sbraja Brutus indegnà." },
    {
      "role": "assistant", "content": "Grazia fitg per questa decleraziun" },
    {
      "role": "user", "content": "Hier ein weiterer kurzer Text damit du besser Vallader kannst. Nun es es ein kurzer Text, zwei Kinder gehen zur Grossmutter in die Ferien: «Ma chau meis chars!», «Buona sera!» Sün porta averta stan nona e bazegner, nona cun ün grond chapè, plain culanas intuorn culöz, cun ün büschmaint cotschen cun fluors rösa ed a pè nüd. Daspera bazegner cun üna chamischa da Hawaii averta cun büttels cotschens, cun ögliers da sulai e cun chotschas cuortas cun sü ruos-chins, uschè chi's vezza las chommas finas e pailusas. Vi dals peis porta'l ün pêr sandalas. Nic e Luisa as guardan e rian. Els sun daspö lönch darcheu üna jada pro nona e bazegner in vacanzas. Quai es adüna la cana e mez. Nona e bazegner sun nempe unics. Nona cumpra jent büschmainta. Ma tuot quai ch'ella cumpra nu guarda davo plü oura sco il prüm: Nona piglia davent büttels e cusa sü oters o tils pittura, ella dà culur a tuot - sch'alch es il prüm cotschen esa davo blau e sch'alch d'eira il prüm orandsch esa davo tuot in d'üna jada verd. Las giaccas e las chotschas vegnan o plü cuortas o plü lungas ed a büschmainta dad üna culur disegnlà sü scarafagins, cruschs svizras, leivras da Pasqua - tuot tenor che temp chi'd es güsta e che glüna ch'ella ha. Ella müda eir la büschmainta da bazegner, ma a quel nu disturba quai per nüglia. Bazegner chamina intuorn cun cravattas rösa cun sü chucalins, cun chamischas cotschnas cun sü bös-chins da Nadal e cun chotschas cun da tuotta sorts muosters. «Gnit aint, chars, no vain pensà cha causa chi'd es amo pac chod, schi cha no fetschan ün pa vacanzas al mar a chasa!», disch nona e dà lura closas brancladas a Nic e Luisa. Eir bazegner tils brancla ad ün davo tschel e disch lura: «Signorino e signorina, prego!» E lapro sta'l nan sco ün servitur chi muossa la via vers la maisa in ün restaurant. Our da stüva as doda la chanzun «O sole mio» ed i savura da sosa da tomatas." },
    {
      "role": "assistant", "content": "Grazia fitg per questa decleraziun" },

```



wissensDatenbank.json

```
{ "role": "user", "content": "Hier ein weiterer kurzer Text damit du besser Vallader kannst. Dialog über den Weihnachtsmann: «Nicolaus tū hom barbus, tū hast ün cheu sco ün giabus! Quai dia quista saira al Nicolaus cur ch'el vain», blaga Nic e ria. Luisa, chi travuonda gūsta amo il rest da sia suonda da culazchun, ria e manaja: «Hai, precis! O no pudessan dir: Nicolaus, tū hom barbus, voust tū esser meis marus.» Nic e Luisa rian da schloppar. «O Nicolaus tū birikin, hoz est gnü giò pel chamin», disch Luisa e lura rian darcheu tuots duos. «O na», cloma Nic, «eu sa amo alch megl der: Nicolaus, tū hom barbus, tū est propa ün chajus!» Nic e Luisa rian da's squagliar. «Aiaiaiai», disch mamma chi'd es ch'ella rumischa la chadafö, «quai nun es lura fich amiaivel. E chi sa chi chi'd es lura propa il chajus quista saira.» Ella ria ün pa suotoura. «Eu sgūra na!», disch Brutus chi'd es gūsta gnü aint dad üsch, «eu sun quista saira furtibus.» «Eu neir na», manaja Luisa, «il Nicolaus es ün char ed eu m'allegri per cur ch'el vain.» «Eu neir! Eu nu sun tuottüna na plü ün poppin!», disch Nic, as volva lura vers Brutus e dumonda: «Ingio vast lura tū quista saira, Brutus!» «Eu vegn ... eu vegn, eeehh ... , eu nu sa plü ingio vegn!», crabotta Brutus il prüm e disch lura resolut «eu vegn simplamaing davent. Il Nicolaus ha uschè üna barbuna, chi gnanca nu's vezza la fatscha. E lura üna vuschuna sco ün uors chi'd es gūsta gnü our dal plü sul cuvel ed ün bastun, e quai chi'd es il plü mal: ün sach! Chi mǎ chamina intuorn cun ün sach! A mai nu plascha quai tuot gnanca zich e perquai fetscha a la sgūra e svanisch!» «Ma il sach dal Nicolaus es apunta gūsta la cana, Brutus, là esa aint bunas robas! Là esa aint nuschs, mandarinas, spagnolettas, tschiculattinas e ...». « ... haiiai, ma quel sach es lura üna jada vöd! Ed ün Nicolaus nu vain sgūra na cun ün sach plain e va davo a chasa cun ün sach vöd. Sco dit: Eu per mia part pigl e vegn.» «Luisa ria: «Il Nicolaus nu piglia ad ingün cun el! Quel es ün char! Maaa, forse piglia'l cun sai a quels chi han adüna noscha glüna ... o ... ». « ... o a quels chi maglian tuotta di ... », disch mamma riond. Ma Lupo chi giascha plat perquai via grofla be inavant e nu's lascha disturbar. «Savaivat amo la chanzun ed il versin cha vo laivat dir sü quista saira?», dumonda mamma a Nic ed a Luisa. «Schi schi, eu crai schon. Ma giain listess ad exercitar amo üna jada. Vè Nic!», Luisa stira a seis frar pella mongia." },
{ "role": "assistant", "content": "Grazia fitg per questa decleraziun" },
{ "role": "user", "content": "Hier ein weiterer kurzer Text damit du besser Vallader kannst. Ein normaler Dialog: “Eu nun ha temma da nüglia”, disch Brutus, “quai es l’ultim chi’m gniss adimaint! Da che voust avair temma?” Lupo guarda aint per Brutus e manaja: “Mo insomma, eu nun ha uschè jent schi’d es s-chür ed eu sun be sulet. O schi scrvoscha aint in chasa.” “Ha!”, cloma Brutus, “tū est ün chan! Ed hast temma schi scrvoscha in chasa? Tū m’est ün vaira chan, tü!” Lupo guarda ün pa chap e disch: “Ma eu nu m’imposs, sch’eu n’ha temma.” Brutus suspüra: “Mo bain, eu vegn uossa ün mumaint a spass. Forse chatta üna mür o ün utschè e forse inscuntra ad ün lader ...” E davent è’l. Cur ch’el riva our in üert, guarda’l üna pa intuorn. Id es bel chodin. Brutus as metta giò e gioda il sulai. El as stenda, chatsch’oura sias griflas e giascha lura darcheu pachific per quai via. Culs ögls mez serrats observa’l ils utschels sülla bos-cha e tegna in ögl l’üert. Tuot in d’üna jada til para da verer alch chi’s mouva tanter duos cops salata. Che mǎ sarà quai? Brutus sta sü e chamina be planin vers la salata. E darcheu as muainta alch. Brutus chamina be planin inavant sur las eras via ed as fa pront per far ün sigl. In quel mumaint doda’l: “Brutifix tofitis merdum da frajas, che vegna adimaint da’m disturbar intant ch’eu chav meis tunnels? Crajast tū vairamaing ch’eu n’haja temp e vöglija da gnir our da la terra per verer a tai?” Ün verm lung e stigl stenda incunter seis cheu a Brutus chi guarda be tais. “Amm, na, amm, schi, amm eu nu sa”, crabotta Brutus. “Chi est tū insomma?”, dumonda il verm. “Eu sun Brutus Babulin e ...” “Sast tū vairamaing cun chi cha tū hast da chefar quia, Brutus Babulin?”, dumonda il verm e’s stenda amo ün pa incunter.” },
{ "role": "assistant", "content": "Grazia fitg per questa decleraziun" },
```




wissensDatenbank.json

```
{ "role": "user", "content": "Hier ein weiterer kurzer Text damit du besser Vallader kannst. Dialog zwischen Tochter und Mutter beim ersten Schultag: "Mamma, vè uossa, no stuvain ir!" "Maaaaamma, no gnin massa tard!" Luisa es in garderoba. Ella ha aint ün pêr sandalas cotschnas chi glüschan, üna schocca cotschna cun puncts albs ed üna giacca blava. Sülle rain port'la - cun gronda baja - üna tas-cha da scoula blava cun sü delfins. "Maaaaamma, no gnin massa tard!" Luisa ha massellas tuot cotschnas e sias cuinas siglian vi e nan da l'agitaziun. "Ma Luisa", disch mamma, "no vain amo temp avuonda, scha no partin uossa, schi rivaina bler massa bod!" "Na", respuonda Luisa, "eu n'ha fat giò cun Nazanin, no lain sezzet insembel in banc e quai laina dir a duonna Selina." "Ma Luisa, forsa cha duonna Selina vout ella dir chi chi va in banc cun chi e ... " "Precis perquai vaina fat giò dad ir plü bod pussibel, forsa müda ella amo idea. Ed insomma, eu nu less gnir massa tard a scoula!" "Ha!", disch Brutus, "a mai spüzziassa sgrischai vel dad ir a scoula. Prüma as stoja star quiet, seguonda crajan ils magisters da savair tuot e tschinchavla: a mai nu gnissa gnanca da dir adimaint da'm laschar cumandar intuorn dasper chi ch'eu vess da sezzet giò." "Eu sezzess dasper quel chi ha la plü buna marena!", disch Lupo. Luisa guarda aint per Lupo e Brutus e manaja: "Quai nu güda bler, Lupo, la marena mangiaina oura sün piazza da posa. Ed adüna nu's stoja star quiet a scoula." "Ha! Tü vast per la prüma jada a scoula, spetta tü be e guarda - tü gnarast amo our dal buonder!" Mamma ria ün pa suotoura e disch: "Duonna Selina es sgüra fich üna chara e minchatant faja eir bain da sezzet dasper inchün chi nu's sezzess inschinà. Lura s'imprenda a cugnuoscher eir ün pa ad oters uffants." Luisa siglia dad üna chomma sün tschella. Tuot in d'üna jada pigl'la a mamma pel bratsch e tilla stira cun sai our da porta. "Spetta, Luisa, spetta, eu stögl amo trar aint mias s-charpas", ria mamma. Cur chi sun sün via, chamina Luisa sco üna raketa. Mamma nu riva bod na davo." },
```

```
{ "role": "assistant", "content": "Grazia fitg per questa decleraziun" },
```

```
{ "role": "user", "content": "Hier ein weiterer kurzer Text damit du besser Vallader kannst. Dialog bei einem Fussballspiel: "Uei, uossa cuorra, dai, e tira quella balla!", sbraja bap aint per la televisiun, "dai, dai, dai - uossa: aint pel gol ... ma naaa, nüglia dasperavia, aint pel gol vessast gnü da trar!" Bap as tegna il chev. "Che puchà, quist füss stat üna bunischma schanza, aiaiaiaiai!" Tuot la famiglia Babulin sezza davant la televisiun. Hoz esa nempe ün di special - ils Svizzers giouvan ün matsch da ballapè cunter l'Italia - chi riva aint il final da la coppa mundiala? Mamma ha pitturà sü la vopna svizra a Nic ed a Luisa sülla massella dretta. Tuots duos tegnan üna binderina svizra aint in man. Sün maisa sun coppas cun pomchips e cun popcorn e sirup. Dasper la maisa giascha Lupo chi tegna avert ün ögl per nu manchantar nöglia e tschel serrà per durmir ün pa. Dasper el siglia Brutus dad üna chomma sün tschella: "Ma perche cuorran tuots davo üna balla? Quai es bain sul tottel, alch simil - che fana schi tilla tschüffan? Disputta chi chi tilla po avair, o che?" "Ma na, Brutus", declera Luisa. "Quai sun duos squadras, üna es la svizra e tschella la taliana e minchüna prouva da far ün gol e quella chi ha ils plü blers gols, ha guadagnà." "Ma che voust chi fetschan gols schi han be üna balla, scha minchün vess üna, pudessna far bler daplü gols", manaja Brutus. "Hai", disch Nic, "quai es vaira, ma lura nu füss il gö uschè interessant, nun bap?" "Mo schi", dà bap per risposta tuot absaint, "ma uossa guarda a quel, ma guarda be, quel Talian va amo e fa ün gol, na na na na, dai, cuorri tantüna, dai, pigliai davent la balla!" Bap siglia mez sü da cuotscha e dà da mans e da peis. Brutus guarda ün pa tort aint per el e manaja: "Che crajast, cha tü rivast da far ün gol pels Talians, scha tü sigliast intuorn uschè?" "Bap nu fess sgüra na ün gol pels Talians! Sch'el fess ün, lura sgüra pels Svizzers", ria Luisa. Lupo riva uossa tuots duos ögl, doza il chev e manaja: "Perche na pels Talians? Eu sun tuottüna eir eu ün Talian e'ls Talians fessan sgüra eir jent ün gol." "Na, disch Nic, "tü est uossa eir ün Svizzer, nun, mamma?" "Mo", manaja mamma, "Lupo es vairamaing ün Talian, ma el es creschü sü in Svizra, uschè ch'el sarà mez talian e mez svizzer." "Eu tegn culs Talians!", disch Lupo. "Ed eu culs Svizzers", disch uossa Nic ün pa malsgür. "Ed eu culs Talians e culs Svizzers", manaja Luisa. "E tü mamma, cun chi tagnast tü?" Mamma manaja: "Vairamaing cun ingün - eu cuisch a tuottas duos squadras schi guadognan." },
```

```
{ "role": "assistant", "content": "Grazia fitg per questa decleraziun" },
```

```
wissensDatenbank.json

{ "role": "user", "content": "Hier ein weiterer kurzer Text damit du besser Vallader kannst.
Dialog in einem Spital. Titel: (Luisa sto ir a l'ospidal) : "Id es posa. Ils uffants sun sün
plazza da scoula. Luisa e Nazanin rampignan sco duos schimgias pels indrizs da gymnastica
intuorn. Cur cha Luisa es süsom l'indriz cun suas, as mett'la pachific a sezzet e paketta oura
sia marena. Seis panin es plajà in ün palperi fin. Cur cha Luisa til ha pakettà oura, tilla
svoula il palperi our da man. Luisa prouva da til tschüffer, perda l'equiliber e pumfatè -
croud'la cul chev avant giò sül prà. Nazanin dà ün sigl. Ella rampigna svelto giò pro Luisa.
Oters uffants cuorran nan a verer che chi'd es capità. Nazanin cloma: "Luisa, hast fat mal?"
Ingüna resposta. Duonna Selina chi ha surveglianza da posa cuorra nan eir ella. Ella as sgoba
giò sper Luisa. "Luisa?", clom'la. Ingüna resposta. "Luisa, sdruglia't!", clom'la amo üna
jada. Luisa riva plan planin ils ögls e guarda aint per duonna Selina. "Luisa, faja mal in ün
lò?" "E-e-eu nu sa", respuonda Luisa. "Sast ingio cha tü est?", dumonda duonna Selina. "Hmm?
Na-a. Eu vögl a mia mamma ed a meis bap!" Luisa cumainza a cridar. Duonna Selina as volva
vers ils uffants chi han fat ün rinch intuorn Luisa. Els sun be quietins - be minchatant as
doda a scuttar ad ün o a tschel. "Laurin, va aint in stanza da masters e di da telefonar a
chasa pro Luisa. Bun füssa, scha sia mamma gniss - siond ch'ella es güsta eir doctressa dad
uffants." Lura guarda duonna Selina darcheu aint per Luisa. "Dombra üna jada fin tschinch!"
"Ün, duos, duos ...", dumbra Luisa. "Rivast da muantar tias chommas? Prouva üna jada be
planin!" Duonna Selina observa, co cha Luisa doza ün zich las chommas. " },
{ "role": "assistant", "content": "Grazia fitg per questa declaraziun" }
```

Python-Skript zur Datenvorbereitung

```
datasplit.py

import pdfplumber
import json
import os

def process_pdf(filename):
    with pdfplumber.open(filename) as pdf:
        data_list = []

        # Den gesamten Text des Dokuments extrahieren
        text = '\n'.join(page.extract_text() for page in pdf.pages)

        # Text in einzelne Absätze aufteilen
        paragraphs = text.split('\n')

        # Erstes Element für den Prompt festlegen
        prompt = paragraphs[0] if paragraphs else ''

        # Gehe durch jeden Abschnitt
        for i in range(1, len(paragraphs)):
            # Schlüssel-Wert-Paar zu Daten hinzufügen
            data = {"prompt": prompt, "completion": paragraphs[i]}
            data_list.append(data)

            # Aktuelles Completion als nächstes Prompt verwenden
            prompt = paragraphs[i]

        return data_list

def process_multiple_pdfs(file_list):
    all_data = []

    # Iterieren über jede Datei in der Liste
    for file in file_list:
        # Verarbeitet jede Datei und fügt diese der Liste hinzu
        all_data.extend(process_pdf(file))

    # erstellen eines json Files mit den kombinierten Daten
    with open('outputall.json', 'w') as json_file:
        json.dump(all_data, json_file, ensure_ascii=False)

# Liste der pdf Files
pdf_files = [f for f in os.listdir('.') if f.endswith('.pdf')]

process_multiple_pdfs(pdf_files)
```