

Digitalia - Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

Tk.T Anssi Jääskeläinen

Tutkimuspäällikkö

XAMK / Digitalia

Anssi.jaaskelainen@xamk.fi

Xamk / Digitalia

- Perustettu 2015, Xamk & Kansalliskirjasto
- Virallinen Xamkin tutkimusyksikkö 2023
 - Tällä hetkellä 13 henkilöä
- Digitalia.fi ja digitalia.xamk.fi
- Uusi tutkimusinfra Memory Lab, 2023

“Jos jotain pitää tehdä useammin kuin kerran se pitää automatisoida”

Memory Lab – Laiteinvestoinnit

01.04.2022 – 31.10.2023

Kokonaisbudjetti: 692 000 €

Rahoittaja ja päärahoituslähde: Etelä-Savon maakuntaliitto, EAKR
(Opetus- ja kulttuuriministeriön toimialan investointihanke)

Toteuttaja: Xamk

Vahvuusala: Digitaalinen talous

Tutkimuskeskus: Digitalia

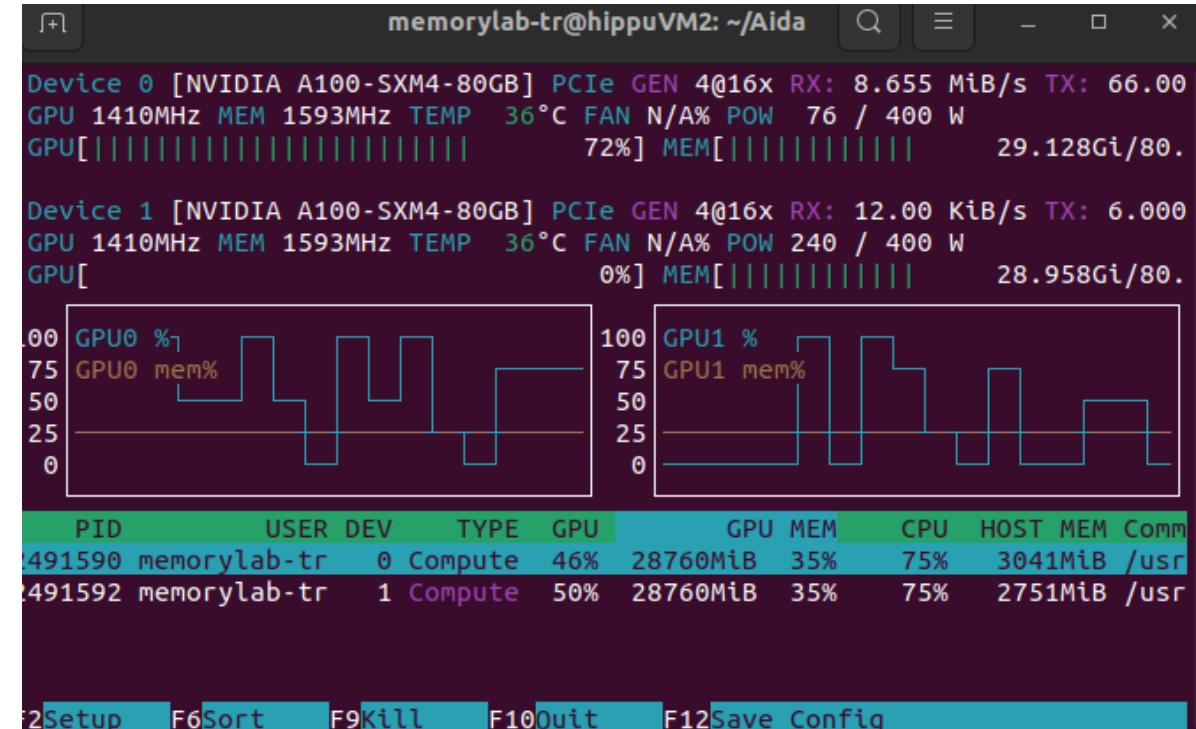
Kestävää kasvua ja työtä -ohjelma

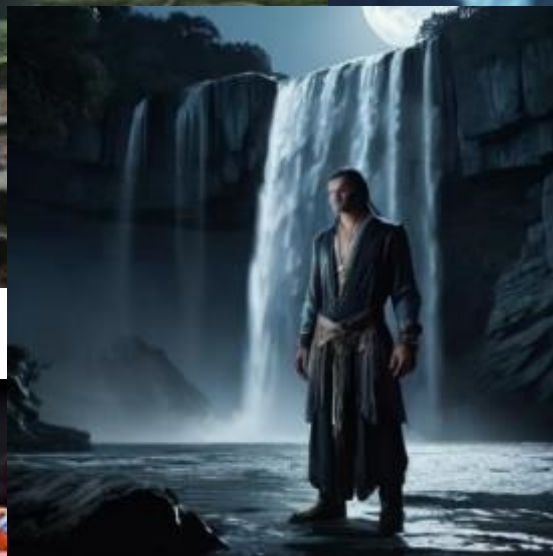
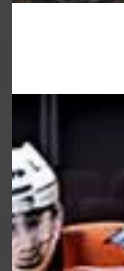
Vipuvoimaa
EU:lta
2014–2020



Kokeiluita kielimalleilla

- Kokeilut HuggingFace & NVIDIA NGC malleilla & päälle tehty oma Python kehitys & Gradio UI
 - <https://memorylab.fi/AIKO/cluetail-demo/>
- OCR luennan tehostaminen
 - <https://memorylab.fi/AIDA/extended-paddle-demo/>
 - Huonosti tunnistetun sanan "arvaus" → a tractor is [Mask] in a field
- Asennettu oma openchat instanssi
 - Auttaa kivasti koodauksessa ja asioiden selittämisessä
- Oma Stable Diffusion asennettuna
 - Generoi kuvia esityksiin yms.





Sähköpostikonversiot: Päivälehti (merkki)

- Entisiä päätoimittajien sähköposteja
 - .pst tiedostossa
 - Työnkulku → PDF/A-3b muotoon metatietoineen + liitetiedostojen erittely ja konversio
 - Osaa .mbox .pst .eml .msg tiedostot
 - Tehtiin myös graafivisualisointi Gephillä
- Hyödyt:
 - Etsintä: indeksoidut pdf tiedostot vs. gigatavuiset .pst tiedostot
 - Arkistokelpoinen formaatti
- Haitat:
 - Ketjut häviävät, jokainen viesti käsitellään omanaan

Automaattinen metatiedotus & muuta

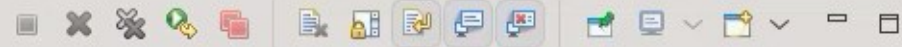
- Dalai –hanke
 - <https://arkkiivi.fi/> ja demo: <http://195.148.220.228/>
 - <https://github.com/DALAI-project>
- Twitter talteenotto
 - Suomen kuntien tilit → sqlite
 - Twitter API
- Python anonymizer
 - NER, sanalistat, jne → Mask
- <https://github.com/Digitalia-Xamk>

Automaattiset käännökset

```
6 from translate import translate
7 from summarize import summarize
8 import time
9 if __name__ == '__main__':
10     translator = translate()
11     summarizer = summarize()
12     starttime = time.time()
13     text = """
14     Majuri Sakari Honkamaan diplomityön LIITE 2 ARVIOINNIN OSUUS KOULUTUKSEN KOKONAISUUDESSA KOULUTUSTARVE TAVOITTEET
15     """
16     translated = translator.Fi_to_Eng(text)[0]["translation_text"]
17     summary = summarizer.summarize(translated)[0]["summary_text"]
18     fi_summary = translator.Eng_to_Fi(summary)[0]["translation_text"]
19     endtime = time.time()
20     runtime = round(endtime - starttime, 2)
21     print("ORG:{}\n TRANS:{}\n SUM:{}\n TRANS SUM:{}\n TIME:{}s".format(
22         text, translated, summary, fi_summary, runtime))
```

time
__main__

Console x



<terminated> MainApp.py [/home/memorylab-aj/anaconda3/envs/aidev/bin/python]

Initializing translate module

Initialize summarizer

ORG:

Majuri Sakari Honkamaan diplomityön LIITE 2 ARVIOINNIN OSUUS KOULUTUKSEN KOKONAISUUDESSA KOULUTUSTARVE TAVOITTEET KOULUTUS OPPIMINEN ASENNE KOULUTUKSEEN MUUTOKSET KÄYTTÄYTYMISESSÄ - Tiedot - Taidot - Asenteet LOPULLINEN TUOTOS - Maanpuolustustahtoinen yksilö - Sotakelpoinen joukko Huom! Arvioinnin osuus on osoitettu katkoviivalla Huom! Kuvio noudattaa pääpiirtein oheista lähdettä. Lähde: Ekola - Vaherva: Aikuisopetusopas 1983, ss. 278-281

TRANS:Major Sakari Honkamaa's Master's thesis ANNEX 2 PROPORTION OF EVALUATION IN THE WHOLENESS OF EDUCATION TRAINING OBJECTIVES TRAINING LEARNING ASSISTANCE TO SCHOOL CHANGES IN USE - Knowledge - Skills - Attitudes FINAL PRODUCTION - National Defence Desired Individual - Warworthy Corps Note! The assessment's contribution has been shown on a dashline Note! The figure follows the main features of the source. Source: Ekola - Vaherva: Adult Education Guide 1983, ps. 278-081

SUM:The assessment's contribution has been shown on a dashline Note! The figure follows the main features of the source. Major Sakari Honkamaa's Master's thesis is shown in its entirety. The assessment's contributions have been shown in their entirety.

TRANS SUM:Arvioinnin panos on esitetty pikavihkossa! Kuvassa seurataan lähteen pääpiirteitä. Majuri Sakari Honkamaan väitöskirja on kokonaisuudessaan esitetty. Arvioinnin tulokset on esitetty kokonaisuudessaan.

TIME:14.93s

Olkaa yhteydessä jos

- Hankeyhteistyö kiinnostaa liittyen
 - Digitaaliseen tiedonhallintaan
 - Tekoälykehitykseen
- Tarvitsette pienimuotoista konsultointia
 - Esim. formaattikonversioissa
 - Tällä hetkellä kunnan kanssa vanhojen Dynasty kantojen muunto käyttökelpoiseen & arkistokelpoiseen muotoon.
 - Avoimen lähdekoodin työkalujen hyödyntämisessä erityisesti arkistoinnin eri vaiheisiin liittyen

