

Laajojen kielimallien ja tietämysgraafien käyttö SampoSammon tietämyksen muodostamisessa

*Annastiina Ahola, väitöskirjatutkija, Aalto-yliopisto
SampoSammon julkistustilaisuus 09.01.2026*

Sisältö

- **Motivaatio**
- **Sääntöpohjainen tiedon poimiminen**
- **Nimettyjen entiteettien tunnistaminen ja linkittäminen**
- **SampoSampo: LLM-pohjainen henkilö–paikka relaatioiden generointi**
- **Muita LLM:ien ja tietämysgraafien yhdistämisen mahdollisuuksia**
- **Lopuksi**

Motivaatio

- **Digitaalisten ihmistieteiden** tutkimuksessa rakenteellinen data (esim. tietämysgraafeista löytyvä data) on tärkeässä roolissa
 - Paljon dataa on kuitenkin saatavilla vain vapaamuotoisena tekstinä (tai kuvamuodossa) + vaihtelevalla määrällä metadataa
- **Laajat kielimallit** (Large Language Model, LLM) tuovat uusia mahdollisuuksia tällaisen datan helpompaan muuntamiseen koneymmärrettävään muotoon
 - Sisällyttämällä prosessissa myös tietämysgraafeja voidaan hyödyntää molempien vahvuuksia (→**neurosymbolinen tekoäly**)
- SampoSammon yhteydessä mielenkiinnon kohteena erityisesti **yhteyksien poimiminen** teksteistä henkilöiden ja paikkojen välillä

Sääntöpohjainen tiedon poimiminen

- Sääntöpohjaista tiedon poimimista voidaan tehokkaasti hyödyntää teksteissä, joissa haluttu poimittava data on ilmaistu säännöllisessä muodossa
 - Esim. synnyin- ja kuolintiedot muodossa
“S [synnyinaika (D)D.(M)M.YYYY-muodossa] [synnyinpaikka perusmuodossa],
K [kuolinaika (D)D.(M)M.YYYY-muodossa] [kuolinpaikka perusmuodossa]”

Ida Emilia Aalberg vsta 1887 Kivekäs vsta 1894 von Uexküll-Gyllenband S 4.12.1857
Janakkala, K 17.1.1915 Pietari. V ratamestari Antti Ahlberg ja Agneta Charlotta Lindroos. P1
1887–1893 Lauri Kivekäs S 1852, K 1893. P1 V kirkkoherra Karl Fr
Ottilia Kristina von Essen; P2 1894–1915 todellinen valtioneuvos, v
Johann von Uexküll-Gyllenband S 1864, K 1923, P2 V salaneuvos,
Uexküll-Gyllenband ja Lina von Adelson.

Ida Aalbergin kansallisbiografia
(<https://biografiasampo.fi/henkilo/p874>)

Yhteyden kuvaus ⓘ	Henkilö	Paikka	Lähde	Yhteyden tyyppi
★ Ida Aalberg on syntynyt paikassa Janakkala vuonna 1857.	Aalberg, Ida (1857 - 1915)	Janakkala	Tapahtuma Semanttisessa kansallisbiografiassa	Syntynyt paikassa
+ Ida Aalberg on kuollut paikassa Pietari vuonna 1915.	Aalberg, Ida (1857 - 1915)	Pietari	Tapahtuma Semanttisessa kansallisbiografiassa	Kuollut paikassa

Nimettyjen entiteettien tunnistaminen ja linkittäminen

- **Nimetty entiteetti:** Entiteetti, johon voi viitata erisnimellä, kuten esimerkiksi henkilö (esim. *Matti Meikäläinen*) tai paikka (esim. *Helsinki*)

PERSON	People, including fictional
NORP	Nationalities or religious or political groups
FACILITY	Buildings, airports, highways, bridges, etc.
ORGANIZATION	Companies, agencies, institutions, etc.
GPE	Countries, cities, states
LOCATION	Non-GPE locations, mountain ranges, bodies of water
PRODUCT	Vehicles, weapons, foods, etc. (Not services)
EVENT	Named hurricanes, battles, wars, sports events, etc.
WORK OF ART	Titles of books, songs, etc.
LAW	Named documents made into laws
LANGUAGE	Any named language

DATE	Absolute or relative dates or periods
TIME	Times smaller than a day
PERCENT	Percentage (including “%”)
MONEY	Monetary values, including unit
QUANTITY	Measurements, as of weight or distance
ORDINAL	“first”, “second”
CARDINAL	Numerals that do not fall under another type

Nimettyjen entiteettien tunnistaminen ja linkittäminen

- **Nimettyjen entiteettien tunnistus (Named Entity Recognition, NER):** Tekstistä tunnistetaan nimetyt entiteetit

“Yritys A:n perustaja Matti Meikäläinen muutti Helsinkiin vuonna 1994.”



NER

“Yritys A:n perustaja Matti Meikäläinen muutti Helsinkiin vuonna 1994.”

(organisaatio)

(henkilö)

(paikka)

(ajankohta)

Nimettyjen entiteettien tunnistaminen ja linkittäminen

- **Nimettyjen entiteettien linkittäminen (Named Entity Linking, NEL):** Nimetyt entiteetit linkitetään yksilöiviin tunnisteihin (esim. Wikidata URI:t)
 - (NER →) kandidaattien generointi → disambiguointi

Esim. onko kirjan tapahtumapaikka Georgia (valtio) vai Georgia (osavaltio, USA)?

“[...] muutti Helsinkiin vuonna [...]”



<http://www.wikidata.org/entity/Q1757>
(Helsinki Wikidatassa)

SampoSampo: LLM-pohjainen henkilö–paikka relaatioiden generointi

- Relaatioiden generoinnin apuna Biografiasammon biografiset tekstit, joille NER + NEL on tehty
→ Henkilön biografiatekstin tunnistetut paikat pohjana mahdollisille relaatioille (jonkinlainen relaatio todennäköisemmin olemassa)

Tunnistettuja paikkoja
(mahdollisia uusia yhteyksiä
Arto Paasilinnalle)

KONTEKSTUAALINEN LUKIJA

Alla olevasta tekstistä on tunnistettu  henkilöt,  paikat, ja  organisaatiot automaattisin menetelmin.

 **Arto Paasilinna** on lukijoiden laajan suosion saavuttanut kirjailija, jonka teokset ovat menestyneet myös ulkomailla. Kirjallisuuskritikoita hänen teoksensa eivät ole onnistuneet vakuuttamaan. Paasilinna on kirjailijana humoristi, ja hänen romaaneidensa aiheet ovat olleet varsin mielikuvituksellisia.

 **Arto Paasilinna** on sodan ja pohjoisen poika. Hänen isänsä työskenteli poliisina  **Petsamossa**, ja Arto syntyi perheen sotapakolaisuuden aikana  **Kittilässä**. Paasilinnan veljessarja on suomalainen menestystarina: Erno Paasilinna oli toisen maailmansodan jälkeisten vuosien kiitetyimpiä, kiistellyimpiä ja palkituimpia kirjailijoita ja  **Reino Paasilinna** aiemmin kohuttu televisio-ohjelmittaja, myöhemmin diplomaatti ja poliitikko. Kirjailijana suurimman kansansuosion on saavuttanut Arto Paasilinna. Hänen tuotantonsa on laaja, valtaosa romaaneja. Lukijoiden suosikkina ja painosten "herrana" Paasilinna sijoittuu  **Kalle Päättälön** ja Laila Hietamiehen (vuodesta 2005 Hirvisaari) rinnalle.

Arto Paasilinnan kansallisbiografia
(<https://biografiasampo.fi/henkilo/p4752>)

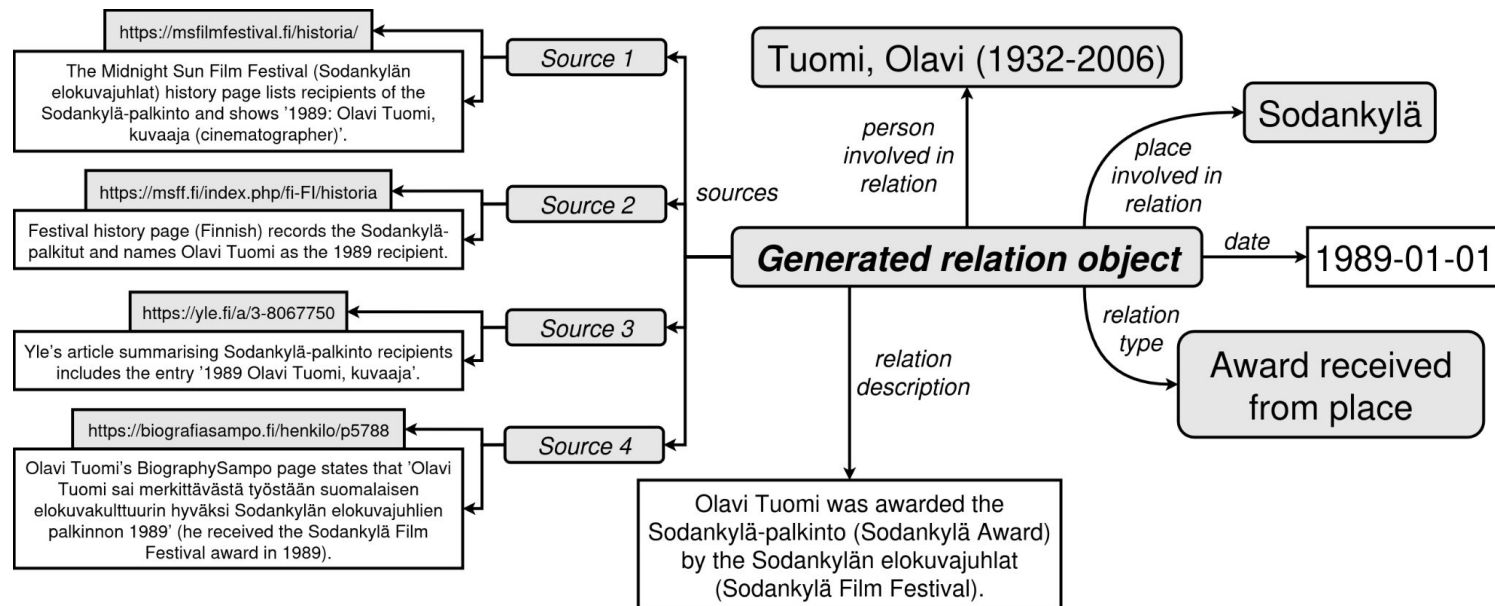
SampoSampo: LLM-pohjainen henkilö–paikka relaatioiden generointi

- **Kielimallin tehtävä yksinkertaistettuna:** Generoi yhteyksiä henkilön *[tietämysgraafista poimittu henkilö]* ja paikan *[valitun henkilön biografiassa mainittu paikka]* välillä hyödyntäen henkilön biografiatekstiä sekä saatavilla olevia internetlähteitä
 - Jokaiselle yhden henkilön biografiassa mainitulle paikalle oma pyyntönsä mallille
 - Joitakin selvästi virheellisesti tunnistettuja paikkoja poistettiin generoitavien henkilö–paikka parien listalta (esim. biografiassa mainittu Johannes hyvin todennäköisesti viittaa henkilöön eikä entiseen kuntaan Johannes)

SampoSampo: LLM-pohjainen henkilö–paikka relaatioiden generointi

- Mallina toimi OpenAI:n **GPT-5 mini** (työkaluina verkkohaku)
- Mallin vastaukset pyydettiin JSON-muodossa, jotka jälkikäteen muutettiin RDF-muotoisiksi ja vastaamaan muiden SampoSampoon sisältyvien yhteyksien datamallia
- Yhteyksien tietojen lisäksi (kuvaus, ajankohta jne.) mallia pyydettiin listaamaan käytetyt internetlähteet ja lyhyt selostus siitä, mitä tietoa malli väittää saaneensa lähteestä
 - Toiveena edes jonkinlainen provenienssitieto, jonka avulla yhteyden todenmukaisuutta voisi alkaa arvioida

SampoSampo: LLM-pohjainen henkilö–paikka relaatioiden generointi



SampoSampo: LLM-pohjainen henkilö-paikka relaatioiden generointi

SampoSampo^{beta}
 Search all content

Person-place relations ⓘ

Results: 939 relations

Active filters:

Data source: Generated using AI ⓘ

REMOVE ALL

Narrow down by:

- Person ⓘ ▾
- Place ⓘ ▾
- Relation type ⓘ ▾

Data source ⓘ

☐ AcademySampo [80306]
☐ BiographySampo [37029]
☐ Wikipedia [26196]
☐ ParliamentSampo [17909]
☐ OperaSampo [16755]
☐ BookSampo [4681]
☐ WarSampo [4562]
☐ ArtSampo [1088]
☐ LetterSampo Finland [997]
☒ Generated using AI [939]
☐ J. V. Snellman Letters [689]
☐ HISTO [331]

Data source of the person ⓘ ▾

TABLE		MAP	SHARE
Person ⓘ	Place ⓘ	Description ⓘ	Relation type ⓘ
Tikkanen, Henrik (1924-1984)	Senaatintori	Henrik Tikkanen depicted the Helsinki square Senaatintori in his visual artwork (ink drawings/cityscapes), with museum-held drawings showing the square and its landmarks.	Generated
Tikkanen, Henrik (1924-1984)	Suomi	Henrik Tikkanen died in Espoo, Suomi (Finland) on 19 May 1984.	Kuoliin paikassa Generated
Tikkanen, Henrik (1924-1984)	Suomi	Henrik Tikkanen worked as a newspaper illustrator / columnist for Finnish newspapers based in Suomi, notably <i>Helsingin Sanomat</i> (1947–1967) and <i>Helsingin Sanomat</i> (1967–1984).	Ura tai opiskelu liittyy paikkaan Generated
Tikkanen, Henrik (1924-1984)	Suomi	Henrik Tikkanen's literary work frequently depicts Helsinki and Finnish social and cultural settings in	

SampoSampo^{beta}
 Search all content

Relation ⓘ

Henrik Tikkanen depicted the Helsinki square Senaatintori in his visual artwork (ink drawings showing the square and its landmarks.)

URI ⓘ	http://idf.fi/reise/r24726333 3418186172-20010574433191
Person ⓘ	Tikkanen, Henrik (1924-1984)
Place ⓘ	Senaatintori
Description ⓘ	Henrik Tikkanen depicted the Helsinki square Senaatintori in drawings showing the square and its landmarks.
Relation type ⓘ	-
Related relations ⓘ	-
Data source ⓘ	Generated using AI
Additional source(s) ⓘ	- BiographySampo states that among Henrik Tikkanen's fav Square), and that he produced numerous cityscapes and i - Helsinki City Museum / Finna catalogue record for a Henr showing Senaatintori in the foreground with people, land


SampoSampo^{beta}



[PEOPLE](#)
[ORGANIZATIONS](#)
[PLACES](#)
[PERSON-PLACE RELATIONS](#)
[DATA SOURCES](#)
[FEED](#)

Relation

Henrik Tikkanen depicted the Helsinki square Senaatintori in his visual artwork (ink drawings/cityscapes), with museum-held drawings showing the square and its landmarks.

TABLE	
URI 	http://idf.fi/relser/r2472633313418186172_20010574433191116739_0
Person 	Tikkanen, Henrik (1924-1984)
Place 	Senaatintori
Description 	Henrik Tikkanen depicted the Helsinki square Senaatintori in his visual artwork (ink drawings/cityscapes), with museum-held drawings showing the square and its landmarks.
Relation type 	-
Related relations 	-
Data source 	Generated using AI
Additional source(s) 	- BiographySampo states that among Henrik Tikkanen's favourite subjects were 'Helsingin Senaatintori' (Helsinki's Senate Square), and that he produced numerous cityscapes and illustrations of Helsinki. - Helsinki City Museum / Finna catalogue record for a Henrik Tikkanen ink drawing (dated to the 1970s) describes the work as showing Senaatintori in the foreground with people, lanterns around the Alexander II statue and the Helsinki Cathedral in the background. - An exhibition description (Henrik Tikkanen's Helsinki) notes that Tikkanen produced thousands of pictures and city reports of Helsinki and in particular depicted the city centre, supporting that his artwork repeatedly represented central Helsinki places such as Senaatintori.

Muita LLM:ien ja tietämysgraafien yhdistämisen mahdollisuuksia

- Tietämysgraafeissa olevia tekstejä voi LLM:ien avulla
 - **asiasanoittaa** ja **klassifioida** haluttuihin kategorioihin
 - poimia **tapahtumia** yhteyksien sijaan
- Tekstien lisäksi LLM:iä voidaan käyttää datan poimimiseen **kuvista** (ja muusta multimodaalisesta datasta valitun mallin tuettujen datatyypin puitteissa), esim. generoimalla asiasanoja tai tekstuaalisia kuvauksia taideteoksista

Lopuksi

- **Laajat kielimallit** mahdollistavat monipuolisemman datan generoinnin ja poiminnan digitaalisten ihmistieteiden aineistoja varten
- **Tietämysgraafien** sisällyttäminen prosessissa parantaa tulosten laatua
 - Tulosten rajaaminen (esim. henkilö–paikka parit SampoSammossa), luotettava konteksti mallille lähteeksi
- **Hallusinaatioista** ei kuitenkaan päästä koskaan kokonaan eroon
→ provenienssitiedon näkyvyys käyttäjälle tärkeää luotettavuuden arvioimista varten



Kiitos!