Extracting Knowledge from Finnish Texts

Digital Humanities in Action: Sampo Model and Portals for Cultural Heritage, 29.10.2020

Minna Tamper

Semantic Computing Research Group (SeCo), Aalto University, *https://seco.cs.aalto.fi*

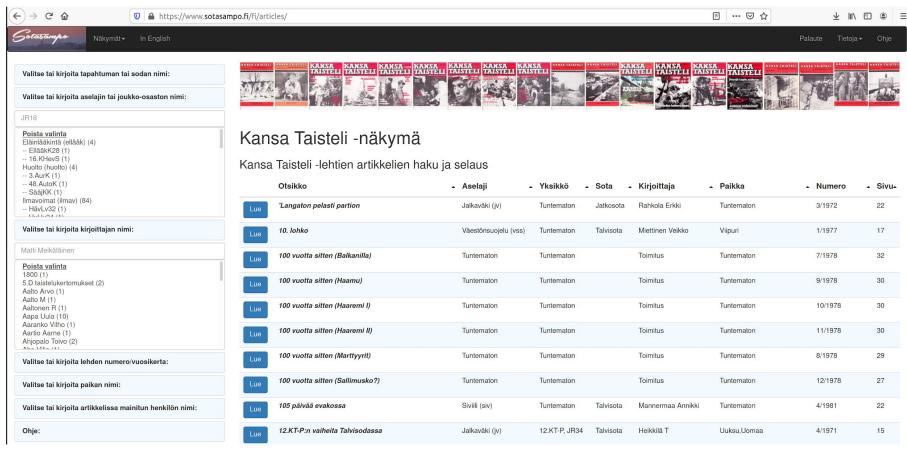*minna.tamper@aalto.fi*

Content
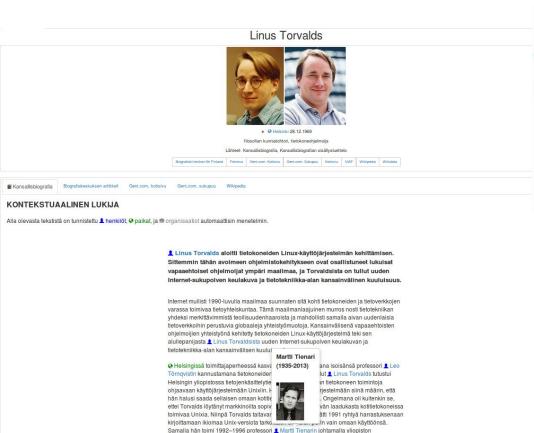
- **Use cases**
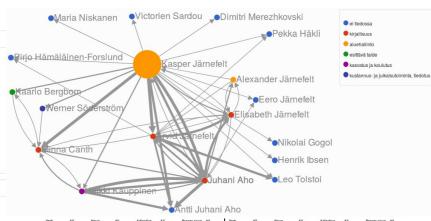- **NLP Services**
- **Questions**

Use cases

WarSampo

BiographySampo

LawSampo

Results

Legend: 👤 person, 🐾 animal, ⚔ mythical or fictional person, 🌐 general location, 📍 address, 🌐 political location (e.g. state), 🌐 geographical location, 🏛 buildings or structures, 📡 astronomical locations (e.g. planets, galaxies), 🏢 organization, 📺 media organization, 💼 financial organizations, 📈 corporation and administration, 📅 date, 🕐 time, 🏷 product, 🚩 event, ⚖ units (e.g. grams, meters) , 💳 money, ▦ registry numbers, social security numbers, 🔨 statutes, case law, ℹ domain information, 📄 title, and 🔨 vocation, and unknown entity ❓.

"ℹ Työntekijän[5] palkkasaatavia koskeva ℹ kanne[3] oli jätetty tutkimatta, koska sitä ei ollut nostettu 🔨 työsopimuslain 13 luvun 9 §[1]:n 3 momentissa säädetyssä kahden vuoden ℹ määräajassa[4] ℹ työsuhteen[2] päättymisestä vaan vasta 📅 3,5[0] vuoden kuluttua siitä. Saman säännöksen mukaan palkkasaatava vanhentuu kuitenkin pykälän 1 momentissa säädetyin tavoin ja siis viiden vuoden kuluttua ℹ erääntymispäivästä[8], jos ℹ työntekijän[5] saatavan perusteena olevia ℹ työehtosopimuksen[7] ℹ määräyksiä[6] on pidettävä ilmeisen tulkinnanvaraisina. ℹ Kanteessa[10] oli ℹ kysymys[9] ℹ työsuhteeseen[2] sovellettavasta ℹ työehtosopimuksesta[7] ja tulkinnanvaraisesta työehtosopimuksen soveltamisalaa koskevasta ℹ määräyksestä[6]. Myös sellaista voitiin pitää palkkasaatavan perusteena olevana ilmeisen tulkinnanvaraisena ℹ määräyksenä[6]. ℹ Kannetta[10] ei olisi saanut jättää 3 momentin nojalla vanhentuneena tutkimatta. 🔨 TSL 13 luku 9 §[11] 3 mom "

📅 **Time**

0. 3,5

🔨 **Statutes, Court decisions**

1. työsopimuslaki 13 luku 9 §
11. TSL 13 luku 9 §

Related documents:

ECLI:FI:KKO:2011:85   ECLI:FI:KKO:2016:19   ECLI:FI:KHO:2001:T894   ECLI:FI:KHO:2001:29   ECLI:FI:KHO:1987:T1261   ECLI:FI:KKO:1980:II77

JSON	Download JSON

NLP Services

Other datasets and services

- **Henko - person name ontology**
  - Based on Finnish Digital Agency's name data and person names collected from other Sampos.
- **Gender identifier**
  - Tool for making a sophisticated guess about person's gener based on Henko's statistics of name usage
- **Person name identifier**
  - Tool to aid in extraction of person names from text, enriching them with other contextual information

Platform for NLP Services

- **Publishing APIs for NLP processing**
  - https://nlp.ldf.fi

SeCo Text Annotation Service is a set of different tools that can be used for Finnish text and language analysis. The tools exchange text to JSON that contains results of each tool.

Select a tool of choice

FiNER
Named entity recognition based on the FIN-CLARIN's FiNER tool.

FinBERT
TurkuNLP's named entity recognition application based on Google's BERT.

APPI - Automatic annotation tool
Named entity linking based on voting scheme.

Finnish dependency parser
Linguistic analysis based on the Turku NLP Finnish-dep-parser

Date and registry number identifiers
Regular expression based application for identifying registry numbers, dates, and more.

Person name finder service
Application that can pick up person names from text and offers links and contextual information (gender, lifespan) about each name.

Gender Identification Service
Application that can determine person's gender based on a name.

Next

- **Service for transforming text to RDF**
- **Applying NLP methods and processes in LawSampo, ParliamentSampo**
- **Application for searching, classifying and exploring legal documents based on semantic content**
- **Other applications coming**

More information

- **Homepage:**
  - https://seco.cs.aalto.fi/projects/nlp/
- **EU COST Action**
  - https://nexuslinguarum.eu/

Questions?