

Teadusartikli struktuur, sihipärane lugemine, kokkuvõtte sõnastamine

DTI6001.DT Õppimine kõrgkoolis

4.10.2019

Loengu kava

- I. Kordamine: publitsistliku teksti kokkuvõte
- II. Teadusteksti kokkuvõtte sisu, struktuur ja märksõnad
- III. Sihipärane lugemine
- IV. Kokkuvõtte koostamine

I. Kordamine: teksti sisukomponendid & teemasõnad;

teksti ülesehituse loogika, struktuur & keel

- Jaaganti poliitikaradar. Paralleelmaailmad (Postimees 19.09.2019), vt

<https://leht.postimees.ee/6782494/>

[jaaganti-poliitikaradar-paralleelmaailmad](#)

- **Rubriik:** Jaaganti poliitikaradar
- **Pealkiri:** Paralleelmaailmad
- **Teema:** Eesti parteide ja poliitikute lahusus (toimitakse paralleelmaailmades)

Teema, fookus, probleem, kokkuvõte

- **Teema:** *koalitsiooni ja opositsiooni vastasseis*
- **Fookus:** *tulised debatid, järjest värvikamad väljendid*
- **Probleem:** *lahendust ja ühist keelt ei paista kumbki pool otsivat; kaevatakse «kruavi» oma poolele*
 - Kirjanduslik paralleel – Pearu ja Andres Tammsaare „Tõest ja õigusest“
- **Kokkuvõte:** tagajärg (tulemus) – toimitakse paralleelselt ehk paralleelmaailmades (kapseldumine)
- **<Lugeja järeldus>**: ei ole jätkusuutlik sisepoliitika, takerdub riigi areng, rahvusvaheline maine saab kannatada

II. Teadusartikli kokkuvõte: tekst

Haridustehnoloogia mõiste sisaldab endas nii õppetöökso sobivaid tehnilisi vahendeid, tarkvara kui ka info- ja kommunikatsioonitehnoloogiaga (IKT-ga) toetatud õpetamistehnoloogiaid. Tehnilised vahendid on õpikeskkonnaks ja tööriistadeks, mis võimaldavad saavutada edukamalt õppe-eesmärke. Selle peatüki eesmärk on tutvustada IKT rolli õpikeskkonna kujundamisel ning meetodeid, kuidas suunata õpilasi kasutama erinevaid vahendeid õpitegevustes (eelkõige arvutit ja mobiiltelefoni) ja toetama oma sotsiaalset arengut. Näidete rõhuasetus on IKT-põhistel õpetamismetoodikatel, mida võiks kasutada põhikooli esimeses ja teises kooliastmes. (Kai Pata 2010. Tehnoloogia. – Õppimine ja õpetamine esimeses ja teises kooliastmes. Tartu: Ecoprint, 417–436.)

Kokkuvõtte sisaldab 4 lauset: esimene lause

- *Haridustehnoloogia mõiste sisaldab endas nii õppetöökõs sobivaid tehnilisi vahendeid, tarkvara kui ka info- ja kommunikatsioonitehnoloogiaga (IKT-ga) toetatud õpetamistehnoloogiaid.*
- **Tutvustab haridustehnoloogia mõistet**
 - **HT (haridustehnoloogia)** —> õppimistehnoloogiad (tarkvara, tehnilised vahendid) —> IKT toetatud õpetamistehnoloogiad
 - [Taust:
 - *õpikeskkonnad*
 - *õpetamismetoodika ja meetodid*]

Teine lause

- *Tehnilised vahendid on õpikeskkonnaks ja tööriistadeks, mis võimaldavad saavutada edukamalt õppe-eesmärke.*

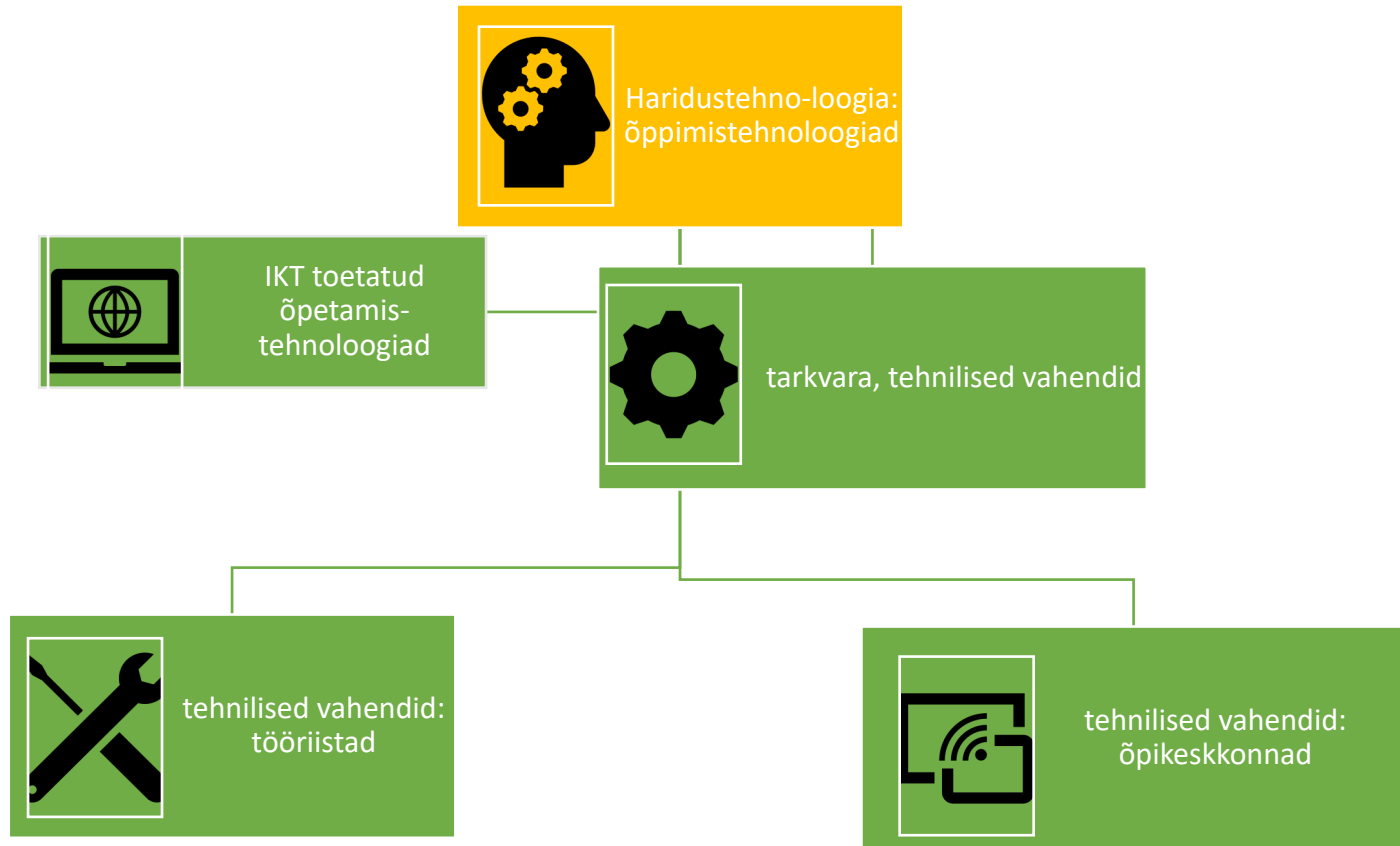
- **Tehniliste vahendite funktsioon**

- olla nii õpikeskkond kui ka tööriistad

[**Taust:** et olla edukam õppe-eesmärkide saavutamisel;

õppimistehnoloogiad (fookuses tehnilised vahendid)]

1. ja 2. lause loogiline struktuur



Kolmas lause

- *Selle peatüki eesmärk on tutvustada IKT rolli õpikeskkonna kujundamisel ning meetodeid, kuidas suunata õpilasi kasutama erinevaid vahendeid õpitegevustes (eelkõige arvutit ja mobiiltelefoni) ja toetama oma sotsiaalset arengut.*
- **Eesmärk:**
 - tutvustada IKT rolli õpikeskkonna kujundamisel;
 - tutvustada meetodeid, kuidas suunata õpilasi kasutama erinevaid vahendeid õpitegevustes ja toetama oma sotsiaalset arengut.

Neljas lause

- *Näidete rõhuasetus on IKT-põhistel õpetamismetoodikatel, mida võiks kasutada põhikooli esimeses ja teises kooliastmes.*
- **Rõhuasetus (fookus):**
 - IKT-põhised õpetamismetoodikad
- **Probleem:**
 - [Taustainfo põhjal: õpiedukuse tõhustamine]

Rõhuasetuse ja probleemi taustad

- **Rõhuasetus (fookus):** IKT-põhised õpetamismetoodikad
 - **Taust:** põhikooli esimene ja teine kooliaste
- **Probleem:** õpiedukuse tõhustamine
 - **Taust:** õpetamismetoodika, kaasaegsed õpikeskkonnad, tehnoloogia toetatud õpitegevused

Kokkuvõtte loogiline struktuur

- **Alusmõistete seletamine:** õppimis- ja õpetamistehnoloogiad > tarkvara ja tehnilised vahendid > tööriistad ja õpikeskkonnad
- **Eesmärgi sõnastamine:** tutvustada IKT rolli õpikeskkonna kujundamisel ja vajalike meetodite leidmisel
 - **Fookus:** IKT põhised õpetamismetoodikad
- **Probleem:** õpiedukuse tõhustamine

Teemasõnad

- HARIDUSTEHNOLLOOGIA
 - õppimistehnoloogiad
 - tehnilised vahendid
 - õpikeskkonnad
 - tööriistad
 - tarkvara
- INFO- JA
KOMMUNIKATSIOONITEHNOLLOOGIAD
 - õpetamistehnoloogiad

Teemasõnad: järg

- ÕPIKESKKONNAD
 - õpetamismetoodikad
 - meetodid
 - õpitegevused



kuidas suunata õpilasi kasutama erinevaid tehnilisi vahendeid erinevates õpitegevustes

Teemad ja taustad

1) Tehnilised vahendid ja õppimise-õpetamise edukus

Taust: põhikooli esimene ja teine kooliaste

2) Tehnoloogia kasutamine põhikooli esimeses ja teises kooliastmes

Taust: tehnilised vahendid kui õpikeskkond ja tööriistad; tarkvara; IKT-põhised õpetamismetoodikad.

Teemad ja taustad: järg

3) IKT toetatud õppimine ja õpetamine

Taust: tehnilised vahendid, tarkvara, IKT-põhised õpetamismetoodikad; põhikooli esimene ja teine kooliaste

Veel teema sõnastusi

- Mis
 - Tehnoloogia toetatud õppimine ja õpetamine
- Mille kasutamine (rakendamine) milles
 - Tehniliste vahendite kasutamine õppeprotsessis
 - Tehnoloogia-põhine õppimine ja õpetamine
- Mille osa (roll, tähtsus, olulisus) milles
 - HT ja IKT osa õppimises ja õpetamises
 - Õppimis- ja õpetamistehnoloogiate roll õppeeesmärkide täitmises jne

Märksõnaloend: näide 1

- Haridustehnoloogia
- Info- ja kommunikatsioonitehnoloogia
- Õppimis- ja õpetamistehnoloogiad
- Õpikeskkonnad
- Õpetamismetoodikad
- Õpetamismeetodid
- Tarkvara ja tööriistad

Märksõnad: näide 2

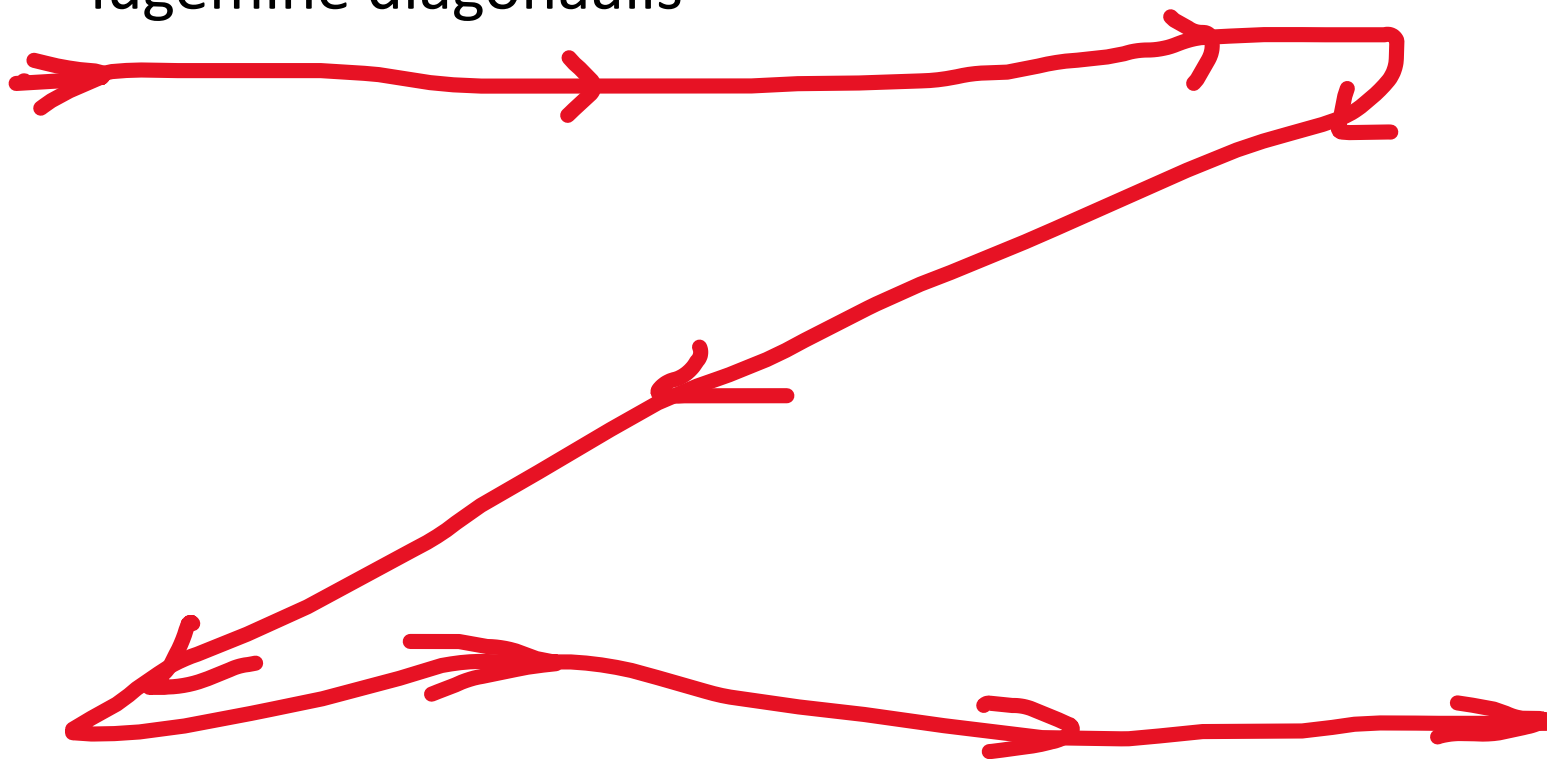
- arvutilingvistika, keeletehnoloogia (*computational linguistics, language technology*)
- arvutilingvistika (*computational linguistics*)
- keeletehnoloogia, kõnetehnoloogia (*language technology, speech technology*)
- keeleressursid (keelevarad) (*language resources*)
- korpuslingvistika, tekstikorpused (*corpus linguistics, linguistic corpora*)
- arvutileksikoloogia, arvutisõnastikud, teaurus (*computational lexicology, computational dictionaries, WordNet(s)*)
- formaalsed grammatikad (*formal grammars*)
- märgendus, süntaktiline (semantiline) märgendus, XML, formaliseerimine (*mark-up, syntactic (semantic) mark-up, encoding, annotation, XML mark-up, formalisation*)

- matemaatiline lingvistika (*mathematical linguistics*)
- formaalsed keeled (*formal languages*)
- automaatne morfoloogia, morfoloogiline ühestamine (*automatic morphology, morphologic disambiguation*)
- keeletöötlus (*natural language processing*)
- süntaktiline töötlus, süntaktiline ühestamine (*parsing, syntactic disambiguation*)
- semantiline töötlus, sõnatähenduste ühestamine (*semantic processing, semantic disambiguation*)
- masintõlge (*machine translation*)

III. Sihipärane lugemine: üks kognitiivne seaduspärasus

Lugeja silmade loomulik liikumine leheküljel kujutab liikumist mööda z-tähe kujutist:

- esmatutvumine tekstiga
- lugemine diagonaalis



Kiirtutvumine allikaga

PEALKIRI

Autor(id): eesnimi + perekonnanimi

Kokkuvõte (annotatsioon, ülevaade)

Võtmesõnad

TEKSTI PÕHIOSA

(Viidatud) kirjandus

Indeksid

Autor(ite) tutvustus koos kontaktandmetega

Ingliskeelne kokkuvõte: autor(id) > kokkuvõtte tekst > võtmesõnad (keywords)

Lisad

Näide 1: Uszkoreit et al. 1997

- <http://www.coli.uni-saarland.de/publikationen/softcopies/Uszkoreit:1997:OFT.pdf> (28.09.2019)
- Kollektiivne monograafia: *Survey of the State of the Art in Human Language Technology*
 - toimetajad: Ron Cole jt
 - Cambridge University Press and Giardini, 1997

Monograafia ülesehitus ehk struktuur

1. Sisukord > missugused teemad, kes on autorid
2. Kolm eessõna > mis on fookuses
k.a monograafia eesmärgi ja struktuuri tutvustus
3. Kas sisaldab infot minu teemaga seotud märksõna(de) või huvipakkuvate autori(te) kohta?

Vt monograafia lõpus

- 1) märksõnaindeksid
nt *natural language processing* (loomuliku keele töötlus)
- 2) viidatud autorid
nt Uszkoreit

Märksõna

natural language processing

510

National Science Foundation of China, 391

natural, 429

natural face, 312

natural language, 155, 211, 227, 238, 245, 299, 303, 420

 dictionaries, 364

 generation, 139, 147, 149, 155, 182, 202, 236, 237

 parsing, 149

 processing, 96, 226, 228, 229, 232, 235, 338, 355, 363, 386, 392, 436

 semantics, 107

 sentences, 111

 software, 239

 system, 205

 technologies, 225

 understanding, 147, 148, 183, 202, 389

natural language processing, 365

natural spoken dialogue, 431

naturally-occurring text, 415, 418

naturally-occurring text corpora, 416

naturalness, 171, 174, 211, 212, 325, 327

nature of discourse relations, 203

nature of discourse segments, 203

nearest neighbor, 65, 344

NEC, 270

necessity, 107

neck, 309

Nestor, 80

network telephony, 327

networking, 329

neural nets, 340

neural network, 6, 76, 81, 273, 310, 324. *see* artificial neural network, 252

NHK, 252

NIST, 273, 389, 402, 437, 451. *see* U.S. National Institute of Standards and Technology

NL, 42–46, 48, 49, 381. *see* natural language

NLG, 139. *see* natural language generation

NLP, 102, 227–229, 234, 360, 361, 384, 394, 409, 411, 418, 419, 427, 428. *see* natural language processing

NN, 24, 381. *see* artificial neural network

noise, 428, 437

 additive, 15, 17, 19

 co-channel interference, 20

 convolutional, 17

 high-intensity, 15

 transient interference, 20

noise removal, 348

noise-suppression, 324

noiseless coding, 329

noisy, 388

noisy characters, 71

noisy environments, 309, 310, 349

noisy measurement, 330

noisy speech, 325

noisy-channel information transmission, 340

nominal accent, 180

non-dynamic logic, 109

non-linear dimension reduction, 357

non-linear documents, 223

non-native speakers, 239

non-parametric voice conversion, 169

noncausal, 332

nonlinear disturbances, 330

INDEX

Start | PDF | File Explorer | Edge | Word | PowerPoint | Chrome | ET | 12:48 | 9.10.2018

Uszkoreit ~ Uzkoreit

Keeleveeb x Survey of the State of the Art in Hu x +

Not secure | www.coli.uni-saarland.de/publikationen/softcopies/Uszkoreit:1997:OFT.pdf

CITATION INDEX 483

Tohkura, Y., 369
Tokuda, L., 89
Tombre, K., 70, 93
Tomita, M., 250, 256,
277, 284, 285
Tomita, Masaru, 114, 130,
133, 137, 355, 379
Tomkins, A., 442
Tong, G., 56
Tong, L. C., 259, 284
Touretzky, D. S., 373, 375,
379, 380
Touretzsky, D. S., 103, 137
Traber, C., 169, 195
Tramus, M. H., 311, 320
Trancoso, I., 390, 408
Trancoso, Isabel, 323
Traum, D. R., 124
Traum, David R., 204, 221
Tremain, T., 323, 336
Trenkle, J. M., 274, 277
Tseng, Gwyneth, 177, 196
Tsujii, J. I., 124, 256,
284
Tsujimoto, S., 68, 93
Tsujimoto, Y., 72, 93
Tsuзaki, Minoru, 169, 191
Tubach, J.P., 196
Turner, R., 110, 137
Tyson, M., 137, 370
Ullman, J. R., 72, 93
Umeda, N., 168, 196
Under, C., 315
Uszkoreit, H., 100, 125,
137, 138, 147, 162
Uszkoreit, Hans, 95, 100,
337
v. Stechow, A., 338, 373
Valbret, H., 174, 196
Valderrama, M. J., 87
van Benthem, J., 132, 313-315,
317
Van Bezooijen, R., 174, 196
Van Compernelle, Dirk, 19, 60,
330, 332, 336
van der Gon, J. J. Denier, 84, 93
van der Horst, K., 441
van Eijck, J., 109, 128,
138
van Emde Boas, P., 132
Van Ess Dykema, C., 281
van Even, S., 276
Van Leeuwen, D.A., 360, 379
van Noord, G., 150, 160,
162
van Noord, Gertjan, 147

ET 12:51 9.10.2018

Monograafia struktuur & ülesehituse loogika

- 13 peatükki, lühendite sõnastik, märksõnaindeks, viidatud autorite indeks
- Peatüki struktuur 12 ptk näitel: **keeleressursid**
 - Sissejuhatav ülevaade
 - Kirjaliku keele korpused
 - Suulise keele korpused
 - Leksikonid
 - Terminoloogia
 - Keelekorpuste nimistu koos linkidega
 - Kirjandus ja indeksid

12. peatükk (keeleressursid): sissejuhatus

- Sissejuhatav ülevaade kaardistab „olukorra riigis“
 - **Antakse mõiste seletus:** „The term **linguistic resources** refers to (usually large) **sets of language data and descriptions in machine readable form, to be used in** building, improving, or evaluating natural language (NL) and speech algorithms or systems“ (Godfrey, Zampolli 1997: 381 jj).
 - **Näited:** „Examples of linguistic resources are **written and spoken corpora, lexical databases, grammars, and terminologies**, although the term may be extended to include **basic software tools for the preparation, collection, management, or use of other resources**“.

12. peatüki sisu, teema aktuaalsus, vajadused

- **Peatüki sisu:** „This chapter deals mainly with **corpora, lexicons, and terminologies**“.
- **Teema aktuaalsus:** „An increasing awareness of the potential economic and social impact of natural language and speech systems has attracted attention, and some support, from national and international funding authorities. **Their interest, naturally, is in technology and systems that work, that make economic sense, and that deal with real language uses** (whether scientifically *interesting* or not)“.
- **Vajadused:** „This interest has been reinforced by the success promised in meeting such goals, by **systems based on statistical modeling techniques such as hidden Markov models (HMM) and neural networks (NN)**, which learn by example, typically from very large data sets organized in terms of many variables with many possible values“.

Probleem(id)

- **Probleem 1:** „A key technical factor in the demand for lexicons and corpora /../ is the enormous appetites of these techniques **for structured data**. **Both** in speech and in natural language, the relatively common occurrence of relatively **uncommon events** (triphones, vocabulary items), and the disruptive effect of even minor **unmodeled events** (channel or microphone differences, new vocabulary items, etc.) means that, **to provide enough examples for statistical methods to work,**
- **Probleem 2:** **the corpora must be numerous** (at the very least one per domain or application), **often massive, and consequently expensive**“. „The **high cost** of creating linguistic resources.“

Rakenduslik tähendus

„ /../ cooperative efforts of companies, research institutions and sponsors /../. This obviously requires that linguistic resources not be restricted to one specific system, but that they be reused—by many users (shareable or public resources), or for more than one purpose (multifunctional resources).“

Tulevikunõuded

- „/.../ linguistic resources must also be *theory-neutral* /.../“
- „/.../ a consensus among different theoretical perspectives and systems design approaches /.../“
- „/.../ the adoption of common specifications and *de facto* standards in creating linguistic resources and ensures their harmonization at the international and multilingual level“
 - The Text Encoding Initiative (TEI) „/.../ has produced a set of guidelines for encoding texts“

Tegevussuunad

- converting existing linguistic resources to common standards; putting some resources in the public domain; to promote the reuse of existing linguistic resources
- to promote the development of new linguistic resources for those languages and domains where they do not exist yet
- to create cooperative infrastructure to collect, maintain, and disseminate linguistic resources on behalf of the research and development community

Lähitulevik

- **Valik konkreetseid tegevusi:**
 - collection of speech and text data
 - creation of Common Lexical Databases with free commercial licenses for members
 - publication of existing corpora previously available only to government contractors
 - intellectual property rights to existing linguistic resources
 - release of government-owned resources to researchers

Tänane seis: nt CLARIN

- CLARIN = Common Language Resources and Technology Infrastructure
 - the mission to create and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences
 - **digital language data** (in written, spoken, or multimodal form) for scholars in the **social sciences** and **humanities**
 - **advanced tools** to discover, explore, exploit, annotate, analyse or combine such data sets, wherever they are located
 - <https://www.clarin.eu/content/services> (CLARINi teenused)

Allika kirotsing

- Teema & märksõnad, autorid & valdkonnad
 - eriakirjandus – tunnustatud teadusajakirjad, raamatukogude andmebaasid, Googel jne > loeng 19. oktoobril
- Leitud tekstid
 - kiirpilk tekstile
 - vaatan tutvustust
 - loen sisukorda
 - leian kokkuvõttest olulised märksõnad
 - kui võimalik, tutvun retsensiooni sisuga, lehitsen lugejate arvamusi

IV. Kokkuvõtte koostamine: sihipärane lugemine

- Mis on **uurimisobjekt** ja valdkond?
- Kas teema tõstatub **objektiivsetest asjaoludest, objekti olemusest**?
- Milles seisneb uurimuse **teoreetilis-rakenduslik olulisus**, milleks on uurimus ja selle tulemused **vajalikud**?
- Kas uuring on/ei ole **aktuaalne**? Miks?
- Milliseid **küsimusi ja hüpoteese** on tõstatatud?
- Mis on uurimuse **eesmärk**?

NB!

eesmärk ei ole uurida, vaid nt seletada, põhjendada, kaardistada, kirjeldada, leida ...

Sihipärane lugemine: järg

- Missugusesse laiemasse ja kitsamasse konteksti uurimus paigutub, mis on uurimuse teoreetiline lähtepunkt?
- Mis **meetodeid, tehnoloogilisi vahendeid ja esituskeelt** on kasutatud?
- Kas uurimistulemused on **originaalsed, kordavad seniseid tulemusi või ei anna midagi uut?**
- Kas tekst on struktureeritud ja kergesti arusaadav, kas teksti sisu esitus on loogiliselt järjestatud?

Kokkuvõtte struktuurikomponendid

- Teema ja autor:
- Uurimisobjekt
- Probleem
- Eesmärk
- Uurimisküsimused
- Materjal & meetodid
- Teoreetiline raamistik
- Analüüsi tulemused
- Diskussioon ja järeldused
- Kokkuvõte

Kokkuvõtte näide: tekstiandmed 1

Survey of the State of the Art in Human Language Technology.
Ed. in Chief Ron Cole. Cambridge (US), 1997. Chapter 12:
Language Resources, 381–384.

Teema: keeleressursid (korpused, leksikonid ja terminoloogiad)

Mõiste seletus: keeleressursi moodustavad keeleliste andmete kogumid ja kirjeldused masinloetaval kujul, mis on mõeldud loomuliku keele ja kõne algoritmide või süsteemide ehitamiseks, arendamiseks või hindamiseks

Teema aktuaalsus: töötavate süsteemide ja tehnoloogia olemasolu, mida saab kasutajaga jagada olenemata sellest, kas sihtrühmal on teadushuvi või rakenduslik vajadus

... tekstiandmed 2

Küsimus: kuidas toimib tegelik keelekasutus ning milleks on need teadmised ühiskonnale vajalikud?

Vajadused: suure ja piiramatu sõnastikuga tuvastussüsteemide tehnoloogia loomine, mis rakendab statistilisi mudeleid, põhiliselt närvivõrke ja/või Markovi varjatud mudeleid (ingl Hidden Markov Models – HMM)

- varjatud Markovi mudel > vene matemaatik Andrei Markov senior

... tekstiandmed 3

Probleemid:

- 1) statistiliste mudelite loomiseks on vaja tohutut hulka struktureeritud keeleandmeid, kuid alati on ebaselgeid sõna-, tähe- või häälikukooslusi ja muid määramata andmeid; see häirib statistilise mudeli efektiivset toimimist ja esitab statistikale väljakutseid
- 2) vaja on tohutu suurt keeleressurssi, kuid keeleressursside loomine on töömahukas ja kallis tegevus

... tekstiandmed 4

Rakenduslik olulisus

1) Ettevõtted, uurimisasutused ja sponsored peaksid tegema koostööd ning keeleressursid peaksid olema **taaskasutatavad, avalikult jagatud, mitmeti kasutatavad**, mitte ainult ühesuguseks analüüsiks sobivad.

2) Üksmeel erinevate teoreetilise-rakenduslike perspektiivide osas.

... tekstiandmed 5

3) Keeleressursside koostamise **universaalsed põhimõtted; teorianeutraalsus.**

4) **Standardite** kasutamine nii rahvusvahelisel tasandil kui ka erinevate keelte suhtes.

5)

... tekstiandmed 6

Tegevussuunad

- 1) Keeleressursside viimine ühtsetele rahvusvahelistele standarditele; ressursside avalik jagamine ja üldsuse teavitamine
- 2) Uute keeleressursside loomine ja propageerimine keeltes, kus neid veel ei ole (sh „väikesed keeled“)
- 3) Koostöövõrgustike loomine, et ressursse säilitada ja levitada.
- 4) Diskussioon tegevussuundade üle erinevates maades.

... tekstiandmed 7

Valik konkreetseid tegevusi

- 1) Suulise kõne ja kirjalike tekstide kogumine
- 2) Leksikaalse andmebaasi loomine (litsentsitasuta)
- 3) Piiratud kasutajaskonnaga ressursside avamine
- 4) Intellektuaalse omandi õiguste kehtestamine ressurssidele
- 5) Riiklike ressursside avamine teadusliku uurimise eesmärgil
- 6) Rahvusvaheline koostöö ressursside loomiseks ja kasutamiseks; vajaliku infrastruktuuri loomine jm

Märksõnade näide

- Keeleressurss
 - Taaskasutus
 - Avatus
- Loomuliku keele töötlus
 - Statistilised mudelid
 - Markovi varjatud mudel
 - Närvivõrgud
- Ressursside võrgustikud
 - Standardus

Kokkuvõtte tekstinäide

Kokkuvõte. Monograafia Survey of the State of the Art in Human Language Technology. Ed. in Chief Ron Cole. Cambridge (US), 1997 peatükk 12 on pühendatud keeleressurssidele, st keeleliste andmete kogudele, mis esitatud masinloetaval kujul ning mõeldud loomuliku keele ja kõne algoritmide (süsteemide) ehitamiseks, arendamiseks või hindamiseks.

Keeleressursside küsimus on aktuaalne, kuna ühiskond vajab töötavaid arvutisüsteeme ja tehnoloogiat, mida saab kasutajaga jagada nii teaduslikel kui ka rakenduslikel eesmärkidel. Seetõttu tuleb ressursse koostades järgida teorianeutraalsust, rakendada universaalseid põhimõtteid ja tehnilisi standardeid.

Tekstitöötuse aluseks on piiramatu sõnastikuga tuvastussüsteemide tehnoloogia loomine, mis rakendab statistilisi mudeleid, põhiliselt närvivõrke ja/või Markovi varjatud mudelit (ingl Hidden Markov Models – HMM).

Monograafia 12. peatükis on kirjeldatud suulise kõne ja kirjalike tekstide ressursse. Peatükk koosneb sissejuhatavast ülevaatest, kirjaliku keele ja suulise kõne korpuste kirjeldustest ning nimistust koos linkidega. Avatud on leksikoni mõiste, seletatud terminoloogiate olulisust. Peatükki lõppu on lisatud kasutatud kirjanduse nimekiri.

Märksõnad: keeleressursid, ressursside võrgustikud, loomuliku keele töötlus, tarkvararakendused, statistilised mudelid, Markovi varjatud mudel, närvivõrgud