

# LÜHENDITE ÜHESTAMINE MEDITSIINITEKSTIDES

2012.-2013. a TIS-i edastatud epikriiside näitel

## PROBLEEM

- Meditsiinitekstides esineb palju dokumenteerimata lühendeid ja akronüüme.
- Lühendile võib vastata mitu täisvormi ja vastupidi.
- See raskendab meditsiinitekstide automaattöötlust.

... käis taastusravis **ca** aasta tagasi ...  
... lisaks vajalik kasutada ka **ca** ja **D-vit** ...  
... **ca** prostatae ...  
... diagnoositud kilpnäärme medullaarne **ca** T4an1aMx

- Lühendi tähendus sõltub tihti kontekstist:
  - ca** võib tähendada kaltsiumi (calcium)
  - ca** võib tähendada kasvajat (carcinoma)
  - ca** võib tähendada hinnangut (circa)

## ANDMED

- Käsitsi koostatud meditsiiniliste lühendite sõnastik:
  - Eola Valdre. Radioloogia terminite sõnastik.
  - Koostamisel vaadati läbi üle 300 000 lühendi.
  - Sisaldab üle 2000 erineva lühendi.
- Epikriisides olevad radioloogiliste uuringute kirjeldused:
  - Kasutada on üle 125 000 uuringu kirjelduse.
  - Neis on üle 50 000 lühendi esinemise.

## TULEMUSED

- Tarkvara lühendite ja akronüümide ühestamiseks.
- Koostamisel on uus radioloogia lühendite sõnastik.
- Tulemusena saab teha täpsemaid teksti otsinguid.

## LÜHENDITE ÜHESTAMISE TÖÖVAHEND

Lühendite ühestamine

Teksti vaade Dokumenti vaade Sõna Indeks Lühendi Indeks

p silm ei näe . kolmas p palavik . vähene p pleurareaktsiooni riba . p 6mm ümargune .

Ühesta lühendid!

Tulemus:

p (parem) silm ei näe . kolmas p (päev) palavik . vähene p (parietaalne) pleurareaktsiooni riba . p (pupill) 6mm ümargune .

Täisvorm	Skoor
pupill	94.47%
parietaalne	5.51%
parem	0.01%

Lühendite ühestamine

Teksti vaade Dokumenti vaade Sõna Indeks Lühendi Indeks

Andmestik: Radioloogia uuringud Text

<< Eelmine Epilid: 89 Järgmine >>

Põhjendus: Öhupuudustunne, hingeldus Kirjeldus: **KI (kliiniline)** andmed: öhupuudustunne, hingeldus Uurituti: kopsud natiivis (7975) ja kontrastainega (7978); mediastinum natiivis (7976) ja kontrastainega (7979); rindkere pehmedkoed natiivis (7976) ja kontrastainega (7979); lülisamba rinnaosa natiivis (7976) ja kontrastainega (7979) Kontrastainena süstitud **Sol (solution)** Visipaque 270 -80,0 ml (milliliiter)

Lühendite ühestamine

Teksti vaade Dokumenti vaade Sõna Indeks Lühendi Indeks

Täisvormi: diagnoos lühendid

Lühend: dg

... hn : Kaelalülid 2s KLIINILISED ANDMED : <b>dg</b> täpsustamiseks LEID : Luude struktuur t ...	100%
... <b>dg</b> valu ja turse vas hüpeliigeses, traum ...	100%
... el ajal sagenenud Obj koldepatoloogiata <b>dg</b> cephalgia Pingepeavalu ? Uuritud peaju ...	100%

Lühend: dgn

... ldi nõeltega 2 x punkteeritud Hemostaas <b>dgn</b> : Lymphadenopathia mediastini ...	100%
... ANDMED : Aprilli alguses dr John poolt <b>dgn</b> -dud vasakul mastit Sai AB-ravi 20.04 ...	100%
... is sin juhtivus blokk sulcuse tasemel, <b>dgn</b> ulnaarkanalil sündroom ( gr m ) ...	100%

Lühend: diagn

... <b>diagn</b> kinnitamiseks Kirjeldus : Luulised stru ...	100%
... lavaljades hingamiskahin karenenud ( Pt <b>diagn</b> ka astma ) Pneumonia sin ? Kirjeldus : ...	100%
... <b>diagn</b> täpsustamiseks Kirjeldus : Lülisamba ni ...	100%

## MEETOD

- Koostame terminite sõnastikust reeglite alusel kõikvõimalikud lühendikandidaadid.
- Iga tekstis esineva lühendi korral järjestame sobivad täisvormid skoori **S** järgi.

$$S(\text{täisvorm}) = P(\text{lühend}|\text{täisvorm}) \times P(\text{kontekst}|\text{täisvorm})$$

- Mudel leiab tasakaalupunkti lühendamistava ja ümbritseva tekstilise konteksti vahel.

