

Potilaskertomus_teksti Miranda ODSissa

Document author: anna.hammais@tyks.fi

Taulut uraodsissa

```
potilaskertomus_ods
potilaskertomus_teksti_ods
```

Linkitys sarakkeella teksti_numero

To XML or not to XML?

Vanhat tekstit ovat XML-muodossa, uudemmat eivät. Uudemmistakin ovat ne, jotka tulevat viestinä muista järjestelmistä kuten QPatista tai Operasta.

Uudenmuotoiset (ei-XML) tekstit

Tila ja historia_avain

Uudenmuotoisilla teksteillä esiintyy seuraavia tiloja: hyväksytty, poistettu, korjattu, muistiinpano. Näistä käytetään vain hyväksytty-tekstejä, muut hylätään. XML:ssä esiintyy lisäksi tila lukittu. XML-teksteistä käytetään niitä, jotka ovat hyväksytty tai lukittu.

Samassa historiassa on saman tekstin eri muokkausversiot. Yleensä historian uusimman tekstin tila on "hyväksytty". Joskus se on "korjattu", jolloin teksti on aina identtinen (teksti_numero on sama eli on vain yksi teksti) kuin sitä edeltävä "hyväksytty" versio. Jos historian uusin teksti on poistettu, samassa historiassa ei ole muita kuin korjattuja tekstejä sitä ennen. On kourallinen historioita, joissa on kaksi hyväksyttyä tekstiä. Nämä saavat molemmat tulla mukaan, koska on ilmeisesti bugi.

Hoitotapahtuma_numero viittaa johonkin seuraavista tauluista:

```
select distinct hoitotapahtuma_tunnus from stage_uraods.potilaskertomus where hoitotapahtuma_tunnus
  is not null;
-- PROSESSITAPAHTUMA
-- OSASTOHOITO
-- VARAUS
-- KÄYNTI
-- LÄHETE
-- JONOVARAUS
```

Hyvin harvoin hoitotapahtuma_numero puuttuu, vaikka teksti onkin uudenmuotoinen.

```
--452
select count(*) from stage_uraods.potilaskertomus as p
inner join stage_uraods.potilaskertomus_teksti as t on p.teksti_numero = t.teksti_numero
where p.hoitotapahtuma_numero is null and teksti !~* '^<';

-- 0
select count(*) from stage_uraods.potilaskertomus as p
inner join stage_uraods.potilaskertomus_teksti as t on p.teksti_numero = t.teksti_numero
where p.hoitotapahtuma_numero is not null and teksti ~* '^<';

# 505 jotka kuuluvat muuhun kuin omaan historiaansa. Nämä ovat uudenmukaisia tekstejä jotka vain
  ilmeisesti bugin johdosta ilman hoitotapahtuma_numeroa.
# Unohdetaan.
# tila      N
# 1: muistiinpano 457
# 2:      korjattu 48
dl[is.na(hoitotapahtuma_numero)][teksti_numero != historia_avain][, .N, by = tila]
```

XML sisältää OID-koodeja, joista voi joskus päätellä, että kyse on esim. väliotsikosta. Alkuosa 1.2.246.537 tarkoittaa Suomea (246), sosiaali- ja terveydenhuoltoa (537). 537:n tilalla 777 tarkoittaa HL7.

Joissain XML:issä on

```
<local_attr name="TAPAHTUMA-AIKA" value="20110306"
<local_attr name="AIKA" value="20110306"
```

, joka kertoo tapahtuma-ajan ja tallennusajan. Kaikissa tällaista ei ole. Näitä parsittu jonkin verran kokeeksi tauluun stage_uraods.tekstit_dom_local_attr.

TAPAHTUMA-AIKA ei aina ole sama kuin XML-rakenteen sisällä vaihtelevissa kohdissa näkyvä päivämäärä. Samoin AIKA ei aina ole sama kuin potilaskertomus-aulun paivityshetki_s, vaikka loogisesti näin luulisi olevan.

Kuitenkin 17M/18M tekstiä menee niin, että TAPAHTUMA-AIKA löytyy local attribuuteista, ja jos löytyy tekstin joukostaikin päivämäärä, se on sama. Noin 700K tekstillä ei ole TAPAHTUMA-AIKAA, ja näistä 17K ei ole AIKAAkaan. Näille käytetään aikaleimana potilaskertomus.paivityshetki_s.

Näyttää siltä, että niillä teksteillä, joiden teksti_numero alkaa numerolla eikä kirjaimella, seuraavista poluista löytyy näkymä, pvm ja yksikkö, jossa palvelutapahtuma on tapahtunut:

```
/section[1]/caption[1] = esim. LKIR (näkymä)
/section[1]/section[2]/paragraph[1]/content[1] = esim. 02.11.2013 (pvm)
/section[1]/section[2]/paragraph[2]/content[1] = esim. LNKIR (yksikkö)
```

Samoin näyttää olevan <section>-alkuisille teksteille (ei aitoa XML:ää.) Nämä saa parsittavaksi kun lisää alkuun XML declarationin eli vaikka

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
```

Näitä parsittu jonkin verran kokeeksi tauluun stage_uraods.temp_dom_alkuosa.

Kirjaimella alkavat teksti_numerot sisältävät sellaisiakin, joilla pvm ja yksikkö löytyvät eri poluista:

```
/section[1]/section[2]/caption[1] = esim. 29.1.03
/section[1]/section[3]/caption[1] = esim. KEU
```

tai

```
/section[1]/section[2]/paragraph[2]/content[1] = pvm
/section[1]/section[2]/paragraph[1]/content[1] = yksikkö
```

Tekstin aikaleimat

Jos tekstillä on potilaskertomus-aulussa hoitotapahtuma_numero, sillä löytyy käynti ja osastohoito, joilla on omat rakenteiset aikaleimansa. Jos hoitotapahtuma_numeroa ei ole, rakenteisena löytyy potilaskertomus.paivityshetki_s, joka on ilmeisesti tekstin päivitysaika Uranuksessa. Sen kuvittelisi olevan sama kuin tekstin sisältä löytyvä local_attr-lohkon "AIKA"-attribuutin aikaleima (6M rivillä se on eri, 11M sama), eli esim. tekstin tallennushetki, mutta nämä aikaleimat ovat usein eri päiviltä. Jos local_attr-lohkosta löytyy "TAPAHTUMA-AIKA", tätä kannattaa käyttää.

Kyselyjä

```
select * from potilaskertomus_teksti_ods where TEKSTI_NUMERO = 'PAT13525958112532';
select * from potilaskertomus_ods where TEKSTI_NUMERO = '256596150928114045';
```

Esimerkki tuoreista teksteistä, jotka eivät ole XML-muodossa:

```
select * from potilaskertomus_teksti_ods t
inner join potilaskertomus_ods k on t.teksti_numero = k.teksti_numero
where extract(year from k.paivitusshetki_s)= 2016
and hoitotapahtuma_numero is not null
and rownum < 1000;
```

Tietty näkymä, tuoreet tekstit:

```
select * from potilaskertomus_ods
where nakyma_selite = 'PAT' and extract(year from paivitusshetki_s) = 2016;
```

Näkymät, joita taulussa esiintyy:

```
select koodiarvo, koodiarvon_selite,
koodiarvo_voimassa_alkaen_pvm, koodiarvo_voimassa_asti_pvm
from stage_uraods.koodisto
where koodiston_aihe = 'URAVI'
order by koodiarvo;
```

Otsikot, joita tekstissä esiintyy:

```
select * from koodisto_ods where koodiston_aihe in ('CLNTI', 'CLLTI');
```

Bugi

Sama teksti_numero liittyy joskus kahteen potilaskertomus_numeroon, joiden molempien tila on hyväksytty tai lukittu. Tällaisia teksti_numeroita on 154 kpl (2016-11-08). Ilmeisesti bugi. Ainakin joissain näistä tapauksista uudempi tekstiversio sisältää tiedon toimipisteestä (potilaskertomus.toimipiste_koodi & nimi) mutta vanha ei. Näin molemmat versiot näkyvät meille, vaikka ovat sama teksti ja toisesta puuttuu toimipiste.

```
select p.teksti_numero,
extract (year from p.paivitusshetki_s), extract(month from p.paivitusshetki_s), count(distinct
p.potilaskertomus_numero)
from potilaskertomus p
inner join potilaskertomus_teksti t on p.teksti_numero = t.teksti_numero
where p.tila in ('lukittu','hyväksytty')
group by p.teksti_numero, extract (year from p.paivitusshetki_s), extract(month from
p.paivitusshetki_s)
having count(distinct p.potilaskertomus_numero) > 1
order by extract (year from p.paivitusshetki_s) desc,
extract(month from p.paivitusshetki_s) desc;
```