



Tiedonhaku potilaskertomuksista ja muusta Ei-rakenteisesta datasta

Kliininen Informatiikka / Toni Mikkola

Kliininen informatiikka

- Olemme osa Tays:n tutkimuspalveluita
- Tarjoamme data- ja analytiikkapalveluita tutkijoille
- Tuotamme mm. Findatan kautta tulleista aineistopyynnöistä potilaskertomuksiin liittyvät aineistopyynnöt

Ei-rakenteinen data potilasjärjestelmässä

- Vapaata tekstiä
 - Erikoissairaanhoidon potilaskertomukset
 - Ns. välilehdet (onk, neur...)
 - Hoitokertomukset, hoito-otsikot ja hoitomerkinnot
 - Laboratorio, patologia, kuvantaminen
 - Pyynnöt, lausunnot, muistiot
 - Muut järjestelmässä olevat kentät, joihin voi vapaasti tekstiä syöttää
 - Esim. diagnoositarkenteet, reseptin annostelu yms.

Potilaskertomustekstien erityispiirteet

- Ensisijaisesti kirjoitettu sairaalan sisäiseen käyttöön muistiin
- Erottuu huomattavasti julkaistuista teksteistä
- Monien eri ammattilaisten kirjoittamia – eri tasoisten kirjoittajien ja eri tyyleillä kirjoitettuja
- Sisältää sensitiivistä henkilötietoa jota voi olla vaikea pseudonymisoida

Potilaskertomustekstien erityispiirteet

- Kirjoitettu usein kiireessä
 - Kirjoitusvirheet, katkenneet lauseet, ei standardit lyhenteet ja akronyymit
- Spekulatiivisten ilmausten ja negaatioiden käyttö yleistä

Milloin hakea muuttujia ei-rakenteisesta datasta

- Haluttu tieto (muuttuja)
 - Ei kirjattu potilastietojärjestelmään rakenteisesti
 - Ei ollenkaan, ei kaikilla osastoilla tai ei koko seuranta-aikaa
 - Ei voida luottaa että rakenteinen kirjaus on tehty
 - Ajanpuutteesta johtuvaa tai kirjaussäännöt ovat muuttuneet ajan mittaan
 - On kirjattu mutta sitä ei erinäisistä syistä ole ketterästi mahdollista saada toisiokäyttöön

Mitä tutkijan kannattaa muistaa tehdessään aineistopyyntöä

- Miettiä kuinka haluttu muuttuja kirjataan kertomukseen
 - Mihin se on kirjattu, kuka ja missä kirjaa ja koska kirjataan
 - Onko se yksiselitteisesti poimittavissa ja helposti muutettavissa muuttujaksi
 - Kuinka se kirjataan: sanavartalot, kirjoitusasut, lyhenteet yms.
 - Luonnollisen kielen epämääräisyys vs standardoitu luokitus

Mitä tutkijan kannattaa muistaa tehdessään aineistopyyntöä

- Potilaskertomuksissa palautettavissa ainoastaan kirjauspäivä koko tekstille, ei välttämättä kerro täyttä totuutta haluttuun muuttujaan liittyvästä ajankohdasta
- Haluttu tieto ei aina välttämättä liity esim. haluttuun diagnoosiin
- Saattaa sisältää hälyä
- Sisältää paljon harmaata aluetta, kun kirjauksia ei löydetä koko kohortille

Mitä tutkijan kannattaa muistaa tehdessään aineistopyyntöä

- Haun tekijä ei ole hoitotyön tekijä ja hän ei välttämättä tiedä kuinka haluttu muuttuja kirjataan ja millä kattavuudella
- Potilastekstin sensitiivisyyden vuoksi pystytään luovuttamaan ainoastaan lyhyitä tekstikatkelmia tai mekaanisesti tunnistettuja lauseita halutun termin ympäriltä
- Jos tutkijalla ei tietoa kuinka haluttu tieto kirjataan sitä voidaan selvittää erillisillä esiselvityksellä

Mitä tutkijan kannattaa muistaa tehdessään aineistopyyntöä

- Hakijan kannattaa suunnitella tarkkaan mihin haku tehdään ja millä termeillä/sanavartaloilla, myös sanallinen kuvaus auttaa haun tekijää ymmärtämään mitä halutaan hakea
- Jos on valmis suunnitelma hausta ja mitä tutkija haluaa aineiston sisältävän, niin se helpottaa kustannusarvion tekemistä ja selkeyttää itse datan keräämistä

Muutama sana aineiston keräykseen käytettävistä työkaluista

- Usein datan keräykseen käytetään säännöllisiä lausekkeita(regular expressions, regex)
 - tällöin myös tutkijalle pystytään kertomaan yksiselitteisesti miksi kyseinen teksti valikoitui mukaan
- Regex palauttaa ainoastaan sen mitä etsitään kaikki muu jää piiloon
 - Usein joudutaan tekemään muutama iteraatio jolla pystytään parantamaan hakutermejä

Muuta

- Tays:lla digitaalisessa muodossa potilaskertomukset vuodesta 2007 eteenpäin



Kysyttävää?



Tays

