

A BNO10 formális formális reprezentálása a GALEN alapján

Surján György^a, Héja Gergely^b

^aEgészségügyi Stratégiai Kutatóintézet

^bBudapest Műszaki és Gazdaságtudományi Egyetem

Áttekintés

- Cél
- BNO10
- openGALEN
- BNO kategóriák modellezése
- Transzformálás OWL-be
- Automatikus klasszifikáció
- Tapasztalatok
- Továbbfejlesztés

Cél

- A BNO10 kategóriáinak formális reprezentálása

Miért?

- Divat
- Diagnózisok automatikus klasszifikálása: DL következtető
- Keresés támogatás népegészségügyi adatbázisokban

BNO10

- Betegségek osztályozó hierarchikus kódrendszer, 5 szinttel
 - legmagasabb szint: főcsoport (pl.fertőző betegségek, légzőrendszer betegségei stb.)
 - legalacsonyabb szint: négy- ill. ötkarakteres tételek, ezek a használható kódok
- Kategória neve
- További információ (kódolási szabályok, szinonimák stb.)

BNO10 problémái

- A betegség finom részletei (pl. szerzett/veleszületett) is befolyásolhatják a klasszifikációt, mely akár több főcsoportot is jelenthet
- “K.m.n.” kategóriák: teljesség
- “Egyéb” / “m.n.o” kategóriák
 - nehéz megtalálni az “egyéb” kategóriákat

openGALEN

- Újrahasznosítható orvosi referencia-ontológia, a fogalmak formálisan definiáltak
- GRAIL: leíró logikaszerű nyelv, de:
 - grammatikai/”értelmes” szintű szankcionálás
 - nincs halmazművelet (\setminus , \cap , \cup)
 - szerep-propagáció

BNO kategóriák ábrázolása

- Főként a megnevezés, de a további információk alapján is
- többdimenziós kódrendszer:
 - anatómia: lokalizáció (ha szükséges)
 - morfológia: elváltozás típusa
 - etiológia: kórok (ha szükséges)
- További specifikáció (diagnózis módja, átvitel módja, szövődmények stb.)

Formális reprezentáció

- (InfectionLesion which <hasSpecificLocation Lung isSpecificConsequenceOf MycobacteriumTuberculosisHominis isDeterminedBy HistologicalTest>) name A15.2
- $A17.8 = A17 \setminus (A17.0 \cup A17.1)$, ez nem írható le GRAIL-ben

Transzformálás OWL-be

- Szükséges, mert:
 - halmazműveletek szükségesek (főként unió és komplement)
 - nyílt forrású DL következtető
- DE:
- Az OWL-ben nincs szerep-propagáció
 - saját OWL interfész
 - szerep-propagáció megkerülése: Pl.
szívbetegség = betegség LocatedIn Szív VAGY
PartOf Szív

Automatikus klasszifikáció

- NLP komponens: statisztikai rendszer tezaurusszal
 - Szövegek mondatokra tördelése “pont-szóköz-nagybetű” szekvencia
 - Mondat szavainak és a tezaurusz szavainak statisztikai hasonlósága
 - Minden szó angolra fordítva (GALEN címke)
 - Anatómia és morfológiai entitások kombinálása betegség-fogalommá
- SILK DL következtető: még fejlesztés alatt

Eredmények

- Kórlapok kézi elemzése:
 - releváns betegségek kinyerése
 - kézi BNO kódolás
- NLP modul:
 - Felidézés: 84%
 - Pontosság: 45% (mivel a rendszer felesleges betegségeket és lokalizációkat talált)

Diszkusszió

- Két főcsoport feldolgozva
- BNO10 problémái:
 - C09 “Mandula rosszindulatú daganata”
C09.0 “Mandulaárok”
 - C86-C90 “Rosszul meghatározott másodlagos és meghatározatlan lokalizációjú rosszindulatú daganatok”
- openGALEN problémák
 - hiányzó fogalmak, temporális és térbeli relációk

Továbbfejlesztés

- Foundational Model of Anatomy alapján:
 - OWL transzformáció
 - formális top-level ontológiával (DOLCE) történő harmonizáció
 - élettant, morfológiát stb. leíró ontológiák hozzáadása

Alkalmazási területek

- Intelligens statisztikai lekérdezések
 - orvosi fogalom megadása
 - a releváns BNO kódok automatikus előállítása
- Számítógéppel támogatott BNO kódolás
- Referencia terminológia szolgáltatás
- BNO "re-engineering"

Támogatók

- eEgészség Program
- IKTA 00126/2002
- IST 507505