TECH_08_A2/2-2008-0092

MODELLALAPÚ SZEMANTIKUS KERESŐ RENDSZER KIDOLGOZÁSA

IDŐKÖZI SZAKMAI BESZÁMOLÓ

2. SZAKASZ

A KONTROLLÁLT NYELV DEFINÍCIÓJA

5.3. melléklet

Alkalmazott Logikai Laboratórium
Szegedi Tudományegyetem

Verziókövetés

dátum	változtatás	szerző
2010-06-20	Vázlat.	Varasdi Károly (ALL)
2010-06-22	1. verzió.	Varasdi Károly (ALL)
2010-12-14	Tagolás, tartalomjegyzék hozzáadása.	Varasdi Károly (ALL)
2011-01-12	Előlap elkészítése, véglegesítés.	Simonyi András (ALL)

Tartalomjegyzék

1. Bevezetés **4**

2. A minimális keresőnyelv **4**

3. Példák **5**

1. Bevezetés

A felhasználó által használt keresőnyelv az angol nyelv egy egyszerűsített változata, amelyben a felhasználó specifikálja az őt érdeklő szemantikai struktúrát, amelyet a rendszer azután azonosít a szabadalmi igénypontokban. A keresőnyelv két változatban dolgozható ki. Az alap az abszolút minimumot jelentő *minimális keresőnyelv*, amelyre a felhasználói visszajelzések alapján egy további *bővítési réteg* építhető.[1]

2. A minimális keresőnyelv

A minimális keresőnyelvi kifejezésre (keresőkifejezésre, *query*-re) vonatkozóan a következő szabályokat mondjuk ki (először csak absztrakt szinten, de a tanulmány végén a bevezetett fogalmak és szabályok alkalmazására látni fogunk konkrét példákat is). A keresőkifejezés egyszerű angol mondatok egy sorozata, amely a következő szabályok megtartásával épül fel.

- A keresőkifejezésben szerepelnek ún. **kulcskifejezések**; ezek olyan, a rendszerbe „behuzalozott” szavak illetve kifejezések, amelyek speciális interpretációval rendelkeznek. A felhasználót arra bátorítjuk, hogy lehetőleg minél több ilyen kulcskifejezést használjon, mert ez jelentősen lecsökkenti a félreértések veszélyét. A kulcskifejezések halmazának meghatározása általános és területfüggő meggondolásoktól függ. A területfüggő kulcskifejezések gyakorisági alapon választhatók ki; esetünkben például a statisztikai vizsgálatra kiválasztott 10000 igénypont gyakorisága alapján.[2]
- Új entitás bevezetése történhet:
  - határozatlan névelős (*a/an*) főnévvel (esetleg összetett főnévvel),
  - névelőtlen főnévvel (összetett főnévvel).
- Már bevezetett entitásra a **said** kulcskifejezéssel hivatkozunk. A felhasználót arra bátorítjuk, hogy minél gyakrabban használja az e kulcskifejezés adta lehetőséget a keresőkérdés megfogalmazásakor.
- Az összetett szavak csoportosítása explicit: ezek első előfordulását a felhasználó [szögletes zárójelek] közé teszi. Ebben a felhasználói felület adhat automatikus segítséget is (pl. főnév-sorozatok nagy valószínűséggel ilyenek, így ezeket a rendszer maga is javasolhatja).
- Tagadást a minimális nyelvben csak két helyen engedünk meg:
  - létige után, és
  - *do/does not* formában főige előtt.

  Beágyazott tagadást[3] a minimális nyelv nem tartalmazhat.

[1] Ez utóbbival ebben a tanulmányban nem foglalkozunk.
[2] Alább a példákban a kulcskifejezéseket **félkövér betűkkel** szedjük.
[3] Például: „Said mixture does not contain a component which is not solvable in alcohol.”

- Felsorolások: a többtagú felsorolások elemei {kapcsos zárójelek között állnak}. A felsorolásoknak két típusa van, amelyek megfelelő belső zárójelezés mellett vegyesen is használhatók:
  - konjunktív: „{NP_1 and NP_2 and ... and NP_n}"
  - diszjunktív: „{NP_1 or NP_2 or ... or NP_n}." A minimális keresőnyelvben diszjunkciót csak itt engedünk meg.
- Valamely entitásról a felhasználó tulajdonságokat a következőképpen mondhat ki:
  - Ha a jellemzett főnév *tárgyi* grammatikai pozícióban van: *which VP* mellékmondati konstrukcióval, akár rekurzívan is (ekkor mindig a legutolsó főnévre vonatkozik a jellemzés). A felhasználót arra bátorítjuk, hogy lehetőleg kerülje az attributív[4] melléknévi módosításokat még a jólhangzás kárára is, hiszen azok gépi elemzése bonyolultabb feladat lehet.
  - Ha a jellemzett főnév *alanyi* pozícióban van:
    - aktív igei állítmánnyal,
    - létige + melléknévi vagy kvantitatív módosítói konstrukcióval, és
    - létige + passzív ige konstrukcióval.

3. Példák

A fentiek alapján a szabadalmi hivatal szakértői által például az US4555399 szabadalomra vonatkozóan küldött keresőkérdések alapján a következő kontrollált keresőnyelvi leírások fogalmazhatók meg.

1. A tablet **contains** aspirin **in the form of** [aspirin crystals]. **Said** aspirin does not irritate the [stomach wall] because **said** aspirin crystals are coated with polymer.
2. A tablet **contains** aspirin **in the form of** [aspirin crystals]. **Said** aspirin crystals are coated with a [polymeric **mixture**] which **constitutes** a [polymeric film] around **said** aspirin crystals. **Said** aspirin interacts with the [gastric fluid] only inside **said** polymeric film.
3. A tablet **contains** aspirin **in the form of** [aspirin crystals]. **Said** aspirin crystals are coated with a [polymeric **mixture**] which **contains** {ethylcellulose and hydroxypropylcellulose}.
4. A tablet **contains** aspirin **in the form of** [aspirin crystals]. **Said** aspirin crystals are coated with a [polymeric **mixture**] which **contains** {1.5 to 15 **parts by weight** ethylcellulose and 1 **part by weight** hydroxypropylcellulose}.

[4] Azaz a főnév fej előtti módosítás. Ellentéte a *predikatív* jellemzés; ez az, amelyet a keresőkifejezésekben használni javasolunk.

5. A tablet **contains** aspirin **in the form of** [aspirin crystals]. **Said** aspirin crystals are 20 to 60 mesh **in size**. **Said** aspirin crystals are coated with a polymer. The weight of **said** polymer is 3 to 10 weight percent.

6. A tablet **contains** aspirin **in the form of** [aspirin crystals]. **Said** aspirin crystals are coated with a [polymeric **mixture**] which **contains** {ethylcellulose and hydroxypropylcellulose and a [compression aid] which is [microcrystalline cellulose] and a disintegrant which is [crossed-linked polyvinylpyrrolidone] and a [tableting aid] which is [hydrogenated vegetable oil]}.